

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2003, 16:59:40 : Search time 93 Seconds
(without alignments)
44.311 Million cell updates/sec

Title: US-09-424-815E-6
Perfect score: 96
Sequence: 1 AKVAKQEKKKKTGRAKRRRA 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	91.7	133	4 Q9H5V4	Q9H5V4 homo sapien
2	88	91.7	133	11 Q9J24	Q9J24 mus musculu
3	88	91.7	137	11 Q920W8	Q920W8 mus musculu
4	88	91.7	137	11 Q920W7	Q920W7 mus spicille
5	88	91.7	137	11 Q91V99	Q91V99 mus musculu
6	83	86.5	132	5 Q9VDH8	Q9VDH8 dtrosophila
7	80	83.5	131	5 Q96201	Q96201 spodoptera
8	76	79.2	133	13 Q90VP1	Q90VP1 ictalurus p
9	65	67.7	130	5 Q18231	Q18231 caenorhabdi
10	64	66.7	130	10 Q9MOE4	Q9MOE4 arabidopsis
11	55	57.3	435	10 Q940Y0	Q940Y0 arabidopsis
12	52	54.2	467	16 Q8R617	Q8R617 t hypocheti
13	50	52.1	350	5 Q9VVF1	Q9VVF1 dtrosophila
14	50	52.1	1051	3 Q01694	Q01694 pneumocysti
15	50	52.1	1630	13 Q90753	Q90753 gallus gall
16	50	52.1	1647	4 Q9HBD3	Q9HBD3 homo sapien

17	50	52.1	1679	4 Q9HBD4	Q9HBD4 homo sapien
18	49	51.0	85	5 Q23154	Q23154 caenorhabdi
19	49	51.0	171	10 Q9LDA1	Q9LDA1 oryza sativ
20	49	51.0	647	5 Q21278	Q21278 caenorhabdi
21	48.5	50.5	224	4 Q96MW1	Q96MW1 homo sapien
22	48	50.0	191	10 Q9L120	Q9L120 oryza sativ
23	48	50.0	207	10 Q80482	Q80482 arabidopsis
24	48	50.0	309	10 Q9S9J0	Q9S9J0 arabidopsis
25	48	50.0	467	10 Q9L182	Q9L182 euglena gra
26	48	50.0	1053	3 Q9HE45	Q9HE45 neurospora
27	47.5	49.5	734	5 Q9VOK7	Q9VOK7 dtrosophila
28	47.5	49.5	749	5 Q95RC1	Q95RC1 dtrosophila
29	47.5	49.5	894	5 Q18462	Q18462 caenorhabdi
30	47	49.0	114	11 Q64561	Q64561 ratu
31	47	49.0	115	11 Q64390	Q64390 ratu
32	47	49.0	117	11 Q91KK4	Q91KK4 mus musculu
33	47	49.0	183	10 Q9ZQ24	Q9ZQ24 arabidopsis
34	47	49.0	196	11 Q9D068	Q9D068 mus musculu
35	47	49.0	257	10 Q8VY29	Q8VY29 arabidopsis
36	47	49.0	414	2 Q93OC7	Q93OC7 bacillus su
37	47	49.0	439	10 Q9SJR3	Q9SJR3 brachydanio
38	47	49.0	510	13 Q90WY5	Q90WY5 streptococ
39	47	49.0	519	2 Q9E2H1	Q9E2H1 streptococ
40	47	49.0	558	10 Q8S3P9	Q8S3P9 oryza sativ
41	47	49.0	649	3 Q94538	Q94538 schizosacch
42	47	49.0	653	11 P97543	P97543 ratu
43	47	49.0	678	4 Q94850	Q94850 homo sapien
44	47	49.0	857	13 Q8UDY6	Q8UDY6 rana pipien
45	46.5	48.4	423	5 Q9NCP4	Q9NCP4 trypanosoma

ALIGNMENTS

RESULT 1

ID	Q9H5V4	PRELIMINARY:	PRT:	133 AA.
AC	Q9H5V4:			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	CNMA: FLJ22986 fis, clone KAT11742.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,			
RA	Suzuki Y., Ohtsushiki M., Nishi T., Shibahara T., Tanaka T.,			
RA	Nakamura Y., Isegal T., Sugano S.,			
RT	"NEBO human cDNA sequencing project."			
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AK026639; BAB15515.1;			
DR	HSSP: P02248; IUBI.			
DR	InterPro: IPR000626; Ubiquitin.			
DR	Pfam: PF00240; ubiquitin; 1.			
DR	PRINTS: PR00348; UBIOUITIN.			
DR	SMART: SM00213; UBO; 1.			
DR	PROSITE: PS00299; UBIOUITIN_1; 1.			
DR	PROSITE: PS0053; UBIOUITIN_2; 1.			
SO	SEQUENCE 133 AA; 14390 MW; 5D2F81F2A355B559 CRC64;			
Query Match 91.7% Score 88; DB 4; Length 133;				
Best Local Similarity 100.0%; Pred. No. 2.3e-05;				
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	2 KVAKQEKKKKTGRAKRR 19			
Db	92 KVAQEKKKKTGRAKRR 109			
RESULT 2				

Q9J24 PRELIMINARY: PRT: 133 AA.
 ID Q9J24
 AC Q9J24: 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DE Monoclonal non-specific suppressor factor beta (Finkel-Biskis-Reilly murine sarcoma virus (FBR-MuSV) ubiquitously expressed) (fox derived).
 DE FAU.
 GN FAU.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SWISS OB:
 RA Nie G.-Y., Li Y., Salomonsen L.A., Clements J.A., Findlay J.K.;
 RT Identification of monoclonal non-specific suppressor factor beta as one of the genes differentially expressed at implantation sites compared to interimplantation sites in the mouse uterus.
 RT Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
 RL [2]
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-C57BL/6J; TISSUE=SMALL INTESTINE, AND KIDNEY;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Aizawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Balarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bull C., Fletcher N., Fujita M., Gariboldi M., Gustinich S., Hill D., Hofmann M., Hume D.A., Kamaya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seiya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S., Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RT Nature 409:685-690(2001).
 RL EMBL: AF147745; AAF60246.1; -;
 DR EMBL: AK008466; BAB25694.1; -;
 DR EMBL: AK002355; BAB22034.1; -;
 DR HSSP: P02248; 10BI.
 DR MGD: MGI:102547; Fau.
 DR InterPro: IPR000626; Ubiquitin.
 DR Pfam: PF00240; ubiquitin.1.
 DR PRINTS: PR00348; UBIQUITIN.
 DR SMART: SM00213; UBO; 1.
 DR PROSITE: PS00299; UBIQUITIN_1; 1.
 DR PROSITE: PS50053; UBIQUITIN_2; 1.
 FT NON_TER
 SQ SEQUENCE 133 AA; 14416 MW; 2087C774A022AB16 CRC64;

Query Match 91.7%; Score 88; DB 11; Length 133;
 Best Local Similarity 100.0%; Pred. No. 2.3e-05;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KVAKEKKKKKTGRAKR 19
 DB 92 KVAKEKKKKKTGRAKR 109

RESULT 3
 ID Q920W8 PRELIMINARY: PRT: 137 AA.
 AC Q920W8: 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Fau protein (Fragment).
 GN FAU.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BL6/MSF;
 RA Liu Y., Kilano T., Koide T., Shirotshi T., Moriwaki K., Saitou N.;
 RT "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of Five Mus musculus subspecies."
 RT Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AB039086; BAB6610.1; -;
 DR InterPro: IPR000626; Ubiquitin.
 DR Pfam: PF00240; ubiquitin.1.
 DR PROSITE: PS00299; UBIQUITIN_1; UNKNOWN_1.
 DR PROSITE: PS50053; UBIQUITIN_2; 1.
 FT NON_TER
 SQ SEQUENCE 137 AA; 14845 MW; 57099FF4055D8B2B CRC64;

Query Match 91.7%; Score 88; DB 11; Length 137;
 Best Local Similarity 100.0%; Pred. No. 2.3e-05;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KVAKEKKKKKTGRAKR 19
 DB 96 KVAKEKKKKKTGRAKR 113

RESULT 4
 ID Q920W7 PRELIMINARY: PRT: 137 AA.
 AC Q920W7: 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Fau protein (Fragment).
 GN FAU.
 OS Mus spicilegus (Steppe mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ZBN;
 RA Liu Y., Kilano T., Koide T., Shirotshi T., Moriwaki K., Saitou N.;
 RT "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of Five Mus musculus subspecies."
 RT Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AB039093; BAB6617.1; -;
 DR InterPro: IPR000626; Ubiquitin.
 DR Pfam: PF00240; ubiquitin.1.
 DR PROSITE: PS00299; UBIQUITIN_1; UNKNOWN_1.
 DR PROSITE: PS50053; UBIQUITIN_2; 1.
 FT NON_TER
 SQ SEQUENCE 137 AA; 14787 MW; 57099FF7065D8B28 CRC64;

Query Match 91.7%; Score 88; DB 11; Length 137;
 Best Local Similarity 100.0%; Pred. No. 2.3e-05;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KVAKEKKKKKTGRAKR 19
 DB 96 KVAKEKKKKKTGRAKR 113

RESULT 5
 ID Q91V99 PRELIMINARY: PRT: 137 AA.
 AC Q91V99:

```

DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Fau protein (Fragment).
CN
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN
RP
RC
RC STRAIN-VARIANTS STRAINS;
RA Liu Y., Kitano T., Koide T., Shiroishi T., Moriwaki K., Saitou N.;
RT "Conspicuous Differences among Gene Genealogies of 21 Nuclear Genes of
RT Five Mus musculus subspecies."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB039084; BAB68608.1; -
DR EMBL; AB039085; BAB68609.1; -
DR EMBL; AB039087; BAB68611.1; -
DR EMBL; AB039088; BAB68612.1; -
DR EMBL; AB039089; BAB68613.1; -
DR EMBL; AB039090; BAB68614.1; -
DR EMBL; AB039091; BAB68615.1; -
DR EMBL; AB039092; BAB68616.1; -
DR InterPro; IPR000626; Ubiqultin.
DR Pfam; PF00240; ubiqultin; 1.
DR PROSITE; PS00299; UBIQUITIN_1; UNKNOWN_1.
DR PROSITE; PSS0053; UBIQUITIN_2; 1.
FT NON_TER
SQ
Query Match 137 AA; 1487 MW; 57099FF7065D8828 CRC64;
Best Local Similarity 100.0%; Pred. No. 2,3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KVAKOEKKKKKTGRAKR 19
DB 96 KVAKOEKKKKKTGRAKR 113

RESULT 6
Q9YVDH8 PRELIMINARY; PRT; 132 AA.
AC Q9YVDH8;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE CG15697 protein (RH08962p).
CN
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN
RP
RC
RC STRAIN-BERKELEY;
RA MEDLINE-20196006; PubMed-10731132;
RX Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanalides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.C., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.F., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Beron P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Bokova D., Botchan M.R., Bouck J., Brokslein P., Brottier P.,
RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

```

```

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Iobagwam C.,
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacled J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA She B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN
RP
RC
RC STRAIN-BERKELEY;
RA Stapleton M., Brokslein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Lao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacled J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celinker S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003732; AAF55815.1; -
DR EMBL; AY071683; AAL49305.1; -
DR Flybase; FBgn0038834; CG15697.
DR InterPro; IPR000626; Ubiqultin.
DR Pfam; PF00240; ubiqultin; 1.
SQ
Query Match 132 AA; 14585 MW; E07C754F3316FDF5 CRC64;
Best Local Similarity 94.4%; Pred. No. 0.00012;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 KVAKOEKKKKKTGRAKR 19
DB 90 KVAKOEKKKKKTGRAKR 107

RESULT 7
Q96201 PRELIMINARY; PRT; 131 AA.
AC Q96201;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Ribosomal protein S30.
OS Spodoptera frugiperda (Fall armyworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Noctuidae; Noctuidae; Amphipyrinae; Spodoptera.
OX NCBI_TaxID=7108;
RN
RP
RC
RC STRAIN-BERKELEY;
RA Landaus I., Ogilastro M., Mita K., Nohata J., Lopez-Ferber M.,
RA Dunon-Cerutti M., Fournier P., Devauchelle G.;
RT "Full-length ribosomal protein sequence from an EST library of
RT Spodoptera frugiperda cells (SF9).";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF400225; AAK92197.1; -
SQ
Query Match 131 AA; 14314 MW; 03AEAD0E31EBC1B04 CRC64;
Best Local Similarity 83.3%; Score 80; DB 5; Length 131;

```


RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Banh J.,
Bowler L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
Ecker J.R.;
RT "Arabidopsis ORF clones";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL161574: CAB79697.1; -;
DR EMBL: AB013392: BAB09885.1; -;
DR EMBL: AY052341: AAK96533.1; -;
DR EMBL: AY061910: AAL1237.1; -;
RW Ribosomal protein.
SQ SEQUENCE 62 AA: 6887 MW: 9508F3EB72F5F33 CRC64;
Query Match 66.7%: Score 64; DB 10; Length 62;
Best Local Similarity 72.2%: Pred. No. 0.033;
Matches 13; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 2 KVAQEKKKKKTGRAKR 19
Db 20 KVAQKKKKKRGRAHKK 37
|||||:|||||:|||||
RESULT 11
ID Q940Y0 PRELIMINARY; PRT; 435 AA.
AC Q940Y0:
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE A15966060/K2A8_13.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shun P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,
Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,
Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,
Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis cDNA clones";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY052363: AAK96554.1; -;
DR InterPro: IPR002110; ANK.
DR Pfam: PF00023; ank; 4.
DR PROSITE: PSS0297; ANK_REPEAT; 2.
DR PROSITE: PSS0297; ANK_REPEAT; 1.
RW ANK repeat; Repeat.
SQ SEQUENCE 435 AA: 48877 MW: DE306D487CC894EC CRC64;
Query Match 57.3%: Score 55; DB 10; Length 435;
Best Local Similarity 64.7%: Pred. No. 3.8;
Matches 11; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

RESULT 12
Q8R617

ID Q8R617 PRELIMINARY; PRT; 467 AA.
AC Q8R617:
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein TTE0033 (Hypothetical protein TTE0744)
DE (Hypothetical protein TTE0838) (Hypothetical protein TTE0852)
DE (Hypothetical protein TTE1082) (Hypothetical protein TTE1247)
DE (Hypothetical protein TTE1519) (Hypothetical protein TTE1678)
DE (Hypothetical protein TTE1739) (Hypothetical protein TTE1823)
DE (Hypothetical protein TTE2212).
GN TTE0033 OR TTE0744 OR TTE0838 OR TTE0852 OR TTE1082 OR TTE1247 OR
GN TTE1519 OR TTE1678 OR TTE1739 OR TTE1823 OR TTE2212.
OS Thermomicrobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermomicrobacteriales; Thermomicrobacteriaceae; Thermomicrobacter.
OX NCBI_Taxid=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MBAT / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002)
DR EMBL: AE012978; AAM23347.1; -;
DR EMBL: AE013042; AAM24005.1; -;
DR EMBL: AE013050; AAM24095.1; -;
DR EMBL: AE013052; AAM24110.1; -;
DR EMBL: AE013072; AAM24327.1; -;
DR EMBL: AE013086; AAM24471.1; -;
DR EMBL: AE013108; AAM24736.1; -;
DR EMBL: AE013122; AAM24879.1; -;
DR EMBL: AE013127; AAM24933.1; -;
DR EMBL: AE013134; AAM25014.1; -;
DR EMBL: AE013165; AAM25364.1; -;
KW Hypothetical protein: Complete proteome.
SQ SEQUENCE 467 AA: 54758 MW: 96B7BA55A25FNA0A CRC64;
Query Match 54.2%: Score 52; DB 16; Length 467;
Best Local Similarity 58.8%: Pred. No. 11;
Matches 10; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 2 KVAQEKKKKKTGRAKR 18
Db 323 KIAEKEKEKIKRAKR 339
|||||:|||||:|||||
RESULT 13
ID Q9VFE1 PRELIMINARY; PRT; 550 AA.
AC Q9VFE1:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE CG13724 protein.
GN CG13724.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
Georgie R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abriil J.F., Agbayan A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Bendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Beran B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Decher A., Deng Z., Kays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreria S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kul P., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mervulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003524; AAF49360.1; -
 DR FlyBase: FBgn0036707; CG13724.
 SQ SEQUENCE 550 AA; 64803 MW; 60524F11BBD4FA64 CRC64;

Query Match 52.1%; Score 50; DB 5; Length 550;
 Best Local Similarity 61.1%; Pred. No. 24;
 Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 2 KVAKOEKKKKTKGRAKR 19
 DB 264 KKYGEKKKKKKGNKRK 281
 ID 001694 PRELIMINARY; PRT; 1051 AA.
 AC 001694;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Glycoprotein A.
 CN GPa.
 OS Pneumocystis carinii.
 OC Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes; Pneumocystidaceae;
 OC Pneumocystis.
 OX NCBI_TaxID=4754;
 RN [1]
 RP SEQUENCE OF 1-191 FROM N.A.
 RC STRAIN-FERRET-DERIVED P. CARINII; TISSUE-INFECTED FERRET LUNG;
 RA Wright T.W., Biscondial T.Y., Haidaris C.G., Gigliotti F.,
 RA Simpson-Haidaris P.J.,
 RT "Isoform Diversity and Tandem Duplication of the Glycoprotein A Gene
 in Ferret Pneumocystis carinii."
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U19871; AAA67766.1; -
 DR InterPro: IPR003330; MSG.
 DR Pfam: PF02349; MSG; 7.
 SQ SEQUENCE 1051 AA; 118853 MW; F587B10DA9B486B9 CRC64;

Query Match 52.1%; Score 50; DB 3; Length 1051;
 Best Local Similarity 50.0%; Pred. No. 44;

Matches 9; Conservative 6; Mismatches 3; Indels 0; Gaps 0;
 OY 1 AKVAKOEKKKKTKGRAKR 18
 DB 762 SKLCKKKKKKKKKDKAKK 779
 ID 090753 PRELIMINARY; PRT; 1630 AA.
 AC 090753;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE BRG1 protein.
 CN BRG1.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HAEMOPOLYTIC;
 RX MEDLINE=97169142; PubMed=9016949;
 RA Goodwin G.H.;
 RT "Isolation of cDNAs encoding chicken homologues of the yeast SNF2 and
 RT Drosophila brhma proteins."
 RL Gene 184:27-32(1997).
 DR EMBL: X91637; CAA62831.1; -
 DR HSP; Q92831; 1891.
 DR InterPro: IPR001487; Bromodomain.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR000330; SNF2_N.
 DR Pfam: PF00439; bromodomain; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00176; SNF2_N; 1.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00297; BROMO; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 DR PROSITE: PS00633; BROMODOMAIN_1; 1.
 DR PROSITE: PS50014; BROMODOMAIN_2; 1.
 DR ATP-binding; Helicase.
 SQ SEQUENCE 1630 AA; 183420 MW; C97CA1DB1DE0DE6C CRC64;

Query Match 52.1%; Score 50; DB 13; Length 1630;
 Best Local Similarity 58.8%; Pred. No. 65;
 Matches 10; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 OY 1 AKVAKOEKKKKTKGRAK 17
 DB 572 AOVAKKKKKKKKKKAE 588

Search completed: February 11, 2003, 17:04:50
 Job time : 96 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2003, 17:01:28 ; Search time 30 Seconds
(without alignments)
19.615 Million cell updates/sec

Title: US-09-424-815E-6
Perfect score: 96
Sequence: 1 AKVAKOEKKKKTKGRAKRA 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2.6/prodata/1/1aa/5A-COMB.pep:*
2: /cgn2.6/prodata/1/1aa/5B-COMB.pep:*
3: /cgn2.6/prodata/1/1aa/6A-COMB.pep:*
4: /cgn2.6/prodata/1/1aa/6B-COMB.pep:*
5: /cgn2.6/prodata/1/1aa/PCTUS-COMB.pep:*
6: /cgn2.6/prodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	52.1	1646	4 US-09-535-008-67	Sequence 67, Appl
2	50	52.1	1647	4 US-09-535-008-2	Sequence 2, Appl
3	50	52.1	1649	4 US-09-535-008-75	Sequence 75, Appl
4	50	52.1	1650	4 US-09-535-008-71	Sequence 71, Appl
5	50	52.1	1678	4 US-09-535-008-69	Sequence 69, Appl
6	50	52.1	1679	4 US-09-535-008-65	Sequence 65, Appl
7	50	52.1	1681	4 US-09-535-008-77	Sequence 77, Appl
8	50	52.1	1684	4 US-09-535-008-73	Sequence 73, Appl
9	47	49.0	22	5 PCT-US95-06077-29	Sequence 29, Appl
10	47	49.0	560	1 US-08-336-618-22	Sequence 22, Appl
11	45	46.9	45	2 US-08-460-890A-58	Sequence 58, Appl
12	45	46.9	45	3 US-08-167-641C-58	Sequence 58, Appl
13	45	46.9	45	4 US-08-460-971A-58	Sequence 58, Appl
14	45	46.9	45	4 US-08-462-040-58	Sequence 58, Appl
15	45	46.9	45	4 US-08-460-890A-60	Sequence 58, Appl
16	45	46.9	59	3 US-08-167-641C-60	Sequence 60, Appl
17	45	46.9	59	3 US-08-460-971A-60	Sequence 60, Appl
18	45	46.9	59	4 US-08-462-040-60	Sequence 60, Appl
19	44	45.8	23	4 US-09-039-780A-9	Sequence 9, Appl
20	44	45.8	1070	4 US-08-697-954-2	Sequence 2, Appl
21	44	45.8	1507	3 US-08-929-329-5	Sequence 5, Appl
22	43	44.8	25	4 US-09-039-780A-118	Sequence 118, Appl
23	43	44.8	63	4 US-08-971-089-2	Sequence 2, Appl
24	43	44.8	128	4 US-09-342-847A-4	Sequence 4, Appl
25	43	44.8	351	4 US-09-245-041-11	Sequence 11, Appl
26	42	43.8	19	4 US-09-039-780A-10	Sequence 10, Appl
27	42	43.8	32	4 US-09-426-680-25	Sequence 25, Appl

28	42	43.8	40	3 US-09-068-650-5	Sequence 5, Appl
29	42	43.8	265	4 US-09-420-592A-5	Sequence 5, Appl
30	42	43.8	269	4 US-09-420-592A-4	Sequence 4, Appl
31	42	43.8	564	1 US-07-872-644-53	Sequence 53, Appl
32	42	43.8	564	1 US-08-297-494-53	Sequence 53, Appl
33	42	43.8	564	1 US-08-297-510-53	Sequence 53, Appl
34	42	43.8	564	1 US-08-479-532-53	Sequence 53, Appl
35	42	43.8	564	1 US-08-455-526-53	Sequence 53, Appl
36	42	43.8	564	1 US-08-455-525-53	Sequence 53, Appl
37	42	43.8	564	3 US-09-139-491-53	Sequence 53, Appl
38	42	43.8	564	5 PCT-US92-03222-53	Sequence 53, Appl
39	42	43.8	634	1 US-07-872-644-51	Sequence 51, Appl
40	42	43.8	634	1 US-08-297-494-51	Sequence 51, Appl
41	42	43.8	634	1 US-08-297-510-51	Sequence 51, Appl
42	42	43.8	634	1 US-08-479-532-51	Sequence 51, Appl
43	42	43.8	634	1 US-08-455-526-51	Sequence 51, Appl
44	42	43.8	634	1 US-08-455-525-51	Sequence 51, Appl
45	42	43.8	634	3 US-09-139-491-51	Sequence 51, Appl

ALIGNMENTS

RESULT 1
US-09-535-008-67
Sequence 67, Application US/09535008
Patent No. 6465629
GENERAL INFORMATION:
APPLICANT: Wong, Alexander K.C.
APPLICANT: Tavligian, Sean V.
TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
TITLE OF INVENTION: AND OTHER CANCER TYPES
FILE REFERENCE: 2318-259
CURRENT APPLICATION NUMBER: US/09/535,008
EARLIER FILING DATE: 2000-03-23
EARLIER APPLICATION NUMBER: U.S. 60/125,806
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 67
LENGTH: 1646
TYPE: PRT
ORGANISM: Homo sapiens
US-09-535-008-67

Query Match 52.1% Score 50: DB 4: Length 1646:
Best Local Similarity 58.8% Pred. No. 42:
Matches 10: Conservative 5: Mismatches 2: Indels 0: Gaps 0:

QY 1 AKVAKOEKKKKTKGRAK 17
I:||||:|||||:|:
DB 574 AOVAKKKKKKKRAE 590

RESULT 2
US-09-535-008-2
Sequence 2, Application US/09535008
Patent No. 6465629
GENERAL INFORMATION:
APPLICANT: Wong, Alexander K.C.
APPLICANT: Tavligian, Sean V.
TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
TITLE OF INVENTION: AND OTHER CANCER TYPES
FILE REFERENCE: 2318-259
CURRENT APPLICATION NUMBER: US/09/535,008
EARLIER FILING DATE: 2000-03-23
EARLIER APPLICATION NUMBER: U.S. 60/125,806
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2

LENGTH: 1647
TYPE: PRT
ORGANISM: Homo sapiens
US-09-535-008-2

Query Match 52.1%; Score 50; DB 4; Length 1647;
Best Local Similarity 58.8%; Pred. No. 42;
Matches 10; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 AKVAKOEKKKKKTGRK 17
1:||||:|||||:|:
DB 574 AQVAKKKKKKKKKAE 590

RESULT 3
US-09-535-008-75
Sequence 75, Application US/09535008
Patent No. 6465629
GENERAL INFORMATION:
APPLICANT: Mong, Alexander K.C.
APPLICANT: Tavligian, Sean V.
APPLICANT: Teng, David H.-F.
TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
FILE REFERENCE: 2318-259
CURRENT APPLICATION NUMBER: US/09/535,008
CURRENT FILING DATE: 2000-03-23
EARLIER APPLICATION NUMBER: U.S. 60/125,806
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 75
LENGTH: 1649
TYPE: PRT
ORGANISM: Homo sapiens
US-09-535-008-75

Query Match 52.1%; Score 50; DB 4; Length 1649;
Best Local Similarity 58.8%; Pred. No. 42;
Matches 10; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 AKVAKOEKKKKKTGRK 17
1:||||:|||||:|:
DB 574 AQVAKKKKKKKKKAE 590

RESULT 4
US-09-535-008-71
Sequence 71, Application US/09535008
Patent No. 6465629
GENERAL INFORMATION:
APPLICANT: Mong, Alexander K.C.
APPLICANT: Tavligian, Sean V.
APPLICANT: Teng, David H.-F.
TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
FILE REFERENCE: 2318-259
CURRENT APPLICATION NUMBER: US/09/535,008
CURRENT FILING DATE: 2000-03-23
EARLIER APPLICATION NUMBER: U.S. 60/125,806
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 71
LENGTH: 1650
TYPE: PRT
ORGANISM: Homo sapiens
US-09-535-008-71

Query Match 52.1%; Score 50; DB 4; Length 1650;
Best Local Similarity 58.8%; Pred. No. 43;
Matches 10; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 AKVAKOEKKKKKTGRK 17
1:||||:|||||:|:
DB 574 AQVAKKKKKKKKKAE 590

RESULT 5
US-09-535-008-69
Sequence 69, Application US/09535008
Patent No. 6465629
GENERAL INFORMATION:
APPLICANT: Mong, Alexander K.C.
APPLICANT: Tavligian, Sean V.
APPLICANT: Teng, David H.-F.
TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
FILE REFERENCE: 2318-259
CURRENT APPLICATION NUMBER: US/09/535,008
CURRENT FILING DATE: 2000-03-23
EARLIER APPLICATION NUMBER: U.S. 60/125,806
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 69
LENGTH: 1678
TYPE: PRT
ORGANISM: Homo sapiens
US-09-535-008-69

Query Match 52.1%; Score 50; DB 4; Length 1678;
Best Local Similarity 58.8%; Pred. No. 43;
Matches 10; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 AKVAKOEKKKKKTGRK 17
1:||||:|||||:|:
DB 574 AQVAKKKKKKKKKAE 590

RESULT 6
US-09-535-008-65
Sequence 65, Application US/09535008
Patent No. 6465629
GENERAL INFORMATION:
APPLICANT: Mong, Alexander K.C.
APPLICANT: Tavligian, Sean V.
APPLICANT: Teng, David H.-F.
TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
FILE REFERENCE: 2318-259
CURRENT APPLICATION NUMBER: US/09/535,008
CURRENT FILING DATE: 2000-03-23
EARLIER APPLICATION NUMBER: U.S. 60/125,806
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 65
LENGTH: 1679
TYPE: PRT
ORGANISM: Homo sapiens
US-09-535-008-65

Query Match 52.1%; Score 50; DB 4; Length 1679;
Best Local Similarity 58.8%; Pred. No. 43;
Matches 10; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 AKVAKOEKKKKKTGRK 17
1:||||:|||||:|:
DB 574 AQVAKKKKKKKKKAE 590

RESULT 7
US-09-535-008-77
Sequence 77, Application US/09535008
Patent No. 6465629


```
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BGI IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; CURRENT FILING DATE: 2000-03-23
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 77
; LENGTH: 1681
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-77

Query Match          52.1%; Score 50; DB 4; Length 1681;
Best Local Similarity 58.8%; Pred. No. 43;
Matches 10; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 AKVAKOEKKKKTKGAK 17
1:||||:|||||:|:
Db 574 AQVAKKKKKKKKRAE 590

RESULT 8
US-09-535-008-73
; Sequence 73, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BGI IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; CURRENT FILING DATE: 2000-03-23
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 73
; LENGTH: 1682
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-73

Query Match          52.1%; Score 50; DB 4; Length 1682;
Best Local Similarity 58.8%; Pred. No. 43;
Matches 10; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 AKVAKOEKKKKTKGAK 17
1:||||:|||||:|:
Db 574 AQVAKKKKKKKKRAE 590

RESULT 9
PCT-US95-06077-29
; Sequence 29, Application PC/TUS9506077
; GENERAL INFORMATION:
; APPLICANT: Immunobiology Research, Institute Inc.
; TITLE OF INVENTION: Vaccine Interdiction of Extracellular
; TITLE OF INVENTION: Transactivating Proteins of Human Immunodeficiency Virus
; TITLE OF INVENTION: And Other Chronically Infecting Viruses Employing Similar
; TITLE OF INVENTION: Intercellular Transactivating Strategies
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESS: Howson and Howson
; STREET: Spring House Corporate Cntr, PO Box 457
```

```
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06077
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/247,991
; FILING DATE: 23-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: RI44PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
PCT-US95-06077-29

Query Match          49.0%; Score 47; DB 5; Length 22;
Best Local Similarity 50.0%; Pred. No. 2;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 2 KVAKOEKKKKTKGRKR 19
1:|:||||:|1111
Db 4 KKRQGRKKRRQGRKKR 21

RESULT 10
US-08-336-618-22
; Sequence 22, Application US/08336618
; Patent No. 5763590
; GENERAL INFORMATION:
; APPLICANT: Peattie, Debra A.
; APPLICANT: Harding, Matthew W.
; APPLICANT: Livingston, David J.
; TITLE OF INVENTION: ISOLATION OF AN MR 52,000 FK506 BINDING
; TITLE OF INVENTION: PROTEIN AND MOLECULAR CLONING OF A CORRESPONDING HUMAN
; TITLE OF INVENTION: CDNA
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESS: Hamilton, Brook, Smith and Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,618
; FILING DATE: 09-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/963,325
; FILING DATE: 16-OCT-1992
; APPLICATION NUMBER: US 07/777,752
; FILING DATE: 11-OCT-1991
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2003, 16:57:29 : Search time 11 seconds

(without alignments)
75.412 Million cell updates/sec

Title: US-09-424-815E-6

Perfect score: 96

Sequence: 1 AKVAKQKKKKKTKGRAKRRA 20

Scoring table: BLOSUM62

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	91.7	59	1	RS30_HUMAN
2	64	66.7	62	1	RS30_ARATH
3	60	62.5	58	1	RS30_PLAFA
4	55	57.3	439	1	AKR_ARATH
5	51	53.1	61	1	RS30_SCHPO
6	50	52.1	59	1	RS30_ORYLA
7	50	52.1	1647	1	SNZ4_HUMAN
8	49.5	51.6	1391	1	MS12_DROHY
9	49	51.0	62	1	RS30_YEAST
10	48	51.0	836	1	UME6_YEAST
11	48	50.0	73	1	H14_RABIT
12	48	50.0	578	1	YBX6_SCHPO
13	48	50.0	862	1	VG01_HSV11
14	48	50.0	1002	1	IF2P_YEAST
15	47	49.0	116	1	STP2_RAT
16	47	49.0	645	1	SUBE_BACSU
17	47	49.0	719	1	DEND_RAT
18	47	49.0	785	1	MUS2_BACHD
19	46	47.9	117	1	STP2_MOUSE
20	46	47.9	131	1	YOHF_BACSU
21	46	47.9	150	1	VPS_BPP2
22	46	47.9	217	1	H1_ANAPL
23	46	47.9	217	1	H1_CHICK
24	46	47.9	218	1	H101_CHICK
25	46	47.9	265	1	RL7A_FUGRU
26	45	46.9	441	1	PLPA_PLACH
27	45	46.9	466	1	CYP8_CAEEL
28	45	46.9	474	1	CBF5_SCHPO
29	45	46.9	529	1	NOP5_HUMAN
30	45	46.9	667	1	YB01_SCHPO
31	45	46.9	708	1	SSRP_MOUSE
32	45	46.9	774	1	GLYB_PLAFA
33	45	46.9	829	1	IF2_HAEIN

34	45	46.9	1064	1	ISK5_HUMAN	09nq38 homo sapien
35	44	45.8	561	1	SSRP_RAT	004931 rattus norv
36	44	45.8	713	1	HS90_EIMTE	044001 elmeria ten
37	44	45.8	2272	1	CCAE_MOUSE	061290 mus musculu
38	43	44.8	132	1	RL7A_PIG	029375 sus scrofa
39	43	44.8	172	1	IF3_HAEIN	P43814 haemophilus
40	43	44.8	188	1	RASK_CYPCA	09yh38 cyprinus ca
41	43	44.8	188	1	RASK_ORYLA	042277 oryzias lat
42	43	44.8	248	1	H1_PARAN	P03256 parecchius
43	43	44.8	265	1	RL7A_HUMAN	P15158 homo sapien
44	43	44.8	265	1	RL7A_ICTPU	090y22 ictalurus p
45	43	44.8	265	1	RL7A_MOUSE	P12970 mus musculu

ALIGNMENTS

```

.. RESULT 1
RS30_HUMAN
AC 005472; Q95261; STANDARD; PRT; 59 AA.
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 40S ribosomal protein S30.
GN
OS Homo sapiens (Human),
OS Mus musculus (Mouse),
OS Rattus norvegicus (Rat),
OS Cricetulus griseus (Chinese hamster), and
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606, 10090, 10116, 10029, 9823;
[1]
RN
RP SPECIES-HUMAN:
RC MEDLINE=92412144; PubMed=1326960;
RA Kas K., Michiels L., Merregaert J.;
RT "Genomic structure and expression of the human fau gene: encoding the
RL ribosomal protein S30 fused to a ubiquitin-like protein.";
RN Biochem. Biophys. Res. Commun. 187:927-933(1992).
[2]
RN
RP SPECIES-HUMAN:
RC MEDLINE=93368957; PubMed=8395683;
RA Michiels L., van der Raaij-van der Wal E., van Hasselt F., Kas K.,
RT "faa cDNA encodes a ubiquitin-like-S30 fusion protein and is
RL expressed as an antisense sequence in the Finkel-Biskis-Kelly murine
RN sarcoma virus.";
RT Oncogene 8:2537-2546(1993).
[3]
RN
RP SPECIES-HUMAN:
RC MEDLINE=95241522; PubMed=7724584;
RA Nakamura M., Xavier R.M., Tsunematsu T., Tanigawa Y.;
RT "Molecular cloning and characterization of a cDNA encoding monoclonal
RN non-specific suppressor factor.";
RT Proc. Natl. Acad. Sci. U.S.A. 92:3463-3467(1995).
[5]
RN
RP SPECIES-HUMAN:
RC MEDLINE=95293388; PubMed=7774934;
RA Casteels D., Pollier C., Guenet J.-L., Merregaert J.;

```

RT "The mouse Fau gene: genomic structure, chromosomal localization, and
RT characterization of two retropseudogenes.";
RL Genomics 25:291-294(1995).
RN [6]
RP SEQUENCE FROM N.A.
RC SPECIES=C. ginseng;
RA Rossman T.G., Wang Z.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RC SPECIES=Pig; TISSUE=Uterus;
RA MEDLINE=97244440; Pubmed=9089280;
RX Chvetzoff S., D'Andrea S.;
RT "Ubiquitin is physiologically induced by interferons in luminal
RT epithelium of porcine endometrium in early pregnancy: global
RT RT-PCR cDNA in place of RNA for differential display screening.";
RL FEBS Lett. 405:148-152(1997).
CC -1- MISCELLANEOUS: THIS RIBOSOMAL PROTEIN IS SYNTHESIZED AS A
CC C-TERMINAL EXTENSION PROTEIN (CEP) OF A UBIQUITIN-LIKE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE 530E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X65921: CAA46714.1: ALT_INIT.
DR EMBL: X65923: CAA46716.1: ALT_INIT.
DR EMBL: X62671: CAA44546.1: -.
DR EMBL: X62671: CAA44545.1: ALT_INIT.
DR EMBL: X65922: CAA46715.1: ALT_INIT.
DR EMBL: D26610: BAA05655.1: ALT_INIT.
DR EMBL: L33715: AAA91564.1: ALT_INIT.
DR EMBL: U41499: AAA83776.1: ALT_INIT.
DR EMBL: U72543: AAB52915.1: ALT_INIT.
DR PIR: JCI278: JCI278.
DR Genew: HGNC:3597: FAU.
DR MIM: 134690: -.
DR MGD: MGI:102547: Fau.
KM Ribosomal protein.
SQ SEQUENCE 59 AA; 6648 MW; 012AC1FB555B01A4 CRC64;
Query Match 91.7%; Score 88; DB 1; Length 59;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 KVAKQKKKKKTGRAKRR 19
DB 18 KVAKQKKKKKTGRAKRR 35
RESULT 2
RS30_ARATH STANDARD: PRT; 62 AA.
ID RS30_ARATH
AC P49689; O82203; O9M0E4;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 40S ribosomal protein S30.
GN (RPS30A OR AT2619750 OR F6P22.22) AND
GN (RPS30B OR AT4629390 OR F17A13.210) AND
GN (RPS30C OR AT5656670 OR M1K19.12).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;

RA Lebas M., Regad F., Lescure B.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (RPS30A).
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; Pubmed=10617197;
RA Lin X., Kaut S., Kounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldlyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pal G., Van Aken S., Unayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhagen G.P., Preuss D.,
RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768(1999).
RN [3]
RP SEQUENCE FROM N.A. (RPS30B).
RC STRAIN=cv. Columbia;
RX MEDLINE=20083488; Pubmed=10617198;
RA Mayer K.F.X., Scheller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Ansgore W., Brandt P., Griwell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaler B., Maché R., Mueller M.,
RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hohenseil J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Baltham L., Robben J.,
RA Van der Schueren J., Glynoprez B., Chuang Y.-J., Vandenbusche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Kert R., Defoor E.,
RA Weltegenner T., Botne G., Kampsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirke W.,
RA Moolman P., Klein Lankhorst R., Rose M., Hauf J., Koelter P.,
RA Bernier S., Hempel S., Feldpausch M., Lambers S., Van den Daele H.,
RA de Keyser A., Buysheert C., Gielen J., Villarroel R., De Clercq R.,
RA van Montagu M., Rogers J., Cronin A., Quail M., Brey-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,
RA Pellet A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Bloecker H., Scharfe M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fartmann B., Granderth K., Danner D., Herzi A.,
RA Neumann S., Argitlou A., Vitale D., Ligouri R., Pivavadi E.,
RA Massenet O., Quigley F., Clabaud G., Muendlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chefor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Bagues M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bietke C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abung-Threideh J.,
RA Stenking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Speth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Droné K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamer B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vili D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
RA Grant S., Shohdy N., Hasegawa A., Hamed A., Lodhi M., Johnson A.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:769-777(1999).
RN [4]
RP SEQUENCE FROM N.A. (RPS30C).
RC STRAIN=cv. Columbia;
RX MEDLINE=98403884; Pubmed=9734815;
RA Kotani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., Miyajima N.,
RA Tabata S.;

RT "Structural analysis of Arabidopsis thaliana chromosome 5. VI.
RT Sequence features of the regions of 1,367,185 bp covered by 19
RT DNA Res. 5:203-216(1998).
RL
RN
RP
RC STRAIN=cv. Columbia;
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "RKEN Arabidopsis full length cDNA clones (RATLs) sequenced by the
RT SSP consortium (Salk/Stanford/PGEC).";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: BELONGS TO THE S30E FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z26869; CAA81482.1; -
DR EMBL; AC005169; AAC62141.2; -
DR EMBL; AL161574; CAB79697.1; -
DR EMBL; AL096692; -; NOT_ANNOTATED_CDS.
DR EMBL; AB013392; BAB09885.1; -
DR EMBL; AY052341; AAK96553.1; -
DR EMBL; AY061910; AAL31237.1; -
RT Ribosomal protein; Multigene family.
FM CONFLICT 17 19 QTP -> RHQ (IN REF. 1).
SQ SEQUENCE 62 AA: 6887 MW: 95DF3EB72F53F33 CRC64;
Query Match
Best Local Similarity 72.7%; Score 64; DB 1; Length 62;
Matches 13; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 2 KVAKOEKKKKTKGRAKR 19
Db 20 KVAKDKKKKKRGRRAHKK 37
IIII:III: III: I

- RESULT 3
RS30_PLAFA STANDARD; PRT: 58 AA.
AC 096269;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 40S ribosomal protein S30.
GN RPS30 OR PFB0885W.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP
RC MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tetteijn H., Carnucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shalton S., Mason T., Ya K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Petlea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum";
RL Science 282:1126-1132(1998).
CC -I- SIMILARITY: BELONGS TO THE S30E FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

CC
DR EMBL; AE001422; AAC71966.1; -
RN Ribosomal protein.
SQ SEQUENCE 58 AA: 6488 MW: 525675AE74D04DE5F CRC64;
Query Match
Best Local Similarity 66.7%; Score 60; DB 1; Length 58;
Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 2 KVAKOEKKKKTKGRAKR 19
Db 20 KVPKDKKKKKRLTGRAKR 37
II: III: III: I

- RESULT 4
AKR_ARATH STANDARD; PRT: 439 AA.
AC 005753;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ankyrin repeat protein (AKRP).
GN AKR.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eustoids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP
RC SEQUENCE FROM N.A.
RC STRAIN=cv. C24; TISSUE=leaf;
RX MEDLINE=93104681; PubMed=1281700;
RA Zhang H., Scheiter D.C., Fowle W.H., Goodman H.M.;
RT "Expression of antisense or sense RNA of an ankyrin repeat-containing
RT gene blocks chloroplast differentiation in Arabidopsis.";
RL Plant Cell 4:1575-1588(1992).
CC -I- FUNCTION: MAY PLAY A ROLE IN THE TEMPORAL AND SPATIAL REGULATION
CC OF CHLOROPLAST DEVELOPMENT FROM PROPLASTID.
CC -I- DEVELOPMENTAL STAGE: HIGHEST EXPRESSION OCCURS IN TWO-WEEK-OLD
CC PLANTS AND DECLINES AS PLANTS DEVELOP FURTHER.
CC
CC -I- SIMILARITY: CONTAINS 4 ANK REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M82863; AAA32812.1; -
DR PIR; J01729; J01729.
DR HSSP; P80144; 2MYO.
DR InterPro: IPR002110; ANK.
DR Pfam: PF00023; ank; 4.
DR SMART: SM00248; ANK; 3.
DR PROSITE: PS50088; ANK_REPEAT; 3.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
KW Cytoskeleton; Repeat; ANK repeat; Multigene family.
FT REPEAT 288 317 ANK 1.
FT REPEAT 321 350 ANK 2.
FT REPEAT 354 383 ANK 3.
FT REPEAT 387 416 ANK 4.
SQ SEQUENCE 439 AA: 49150 MW: C371A90028B2BF3 CRC64;
Query Match
Best Local Similarity 57.3%; Score 55; DB 1; Length 439;
Matches 11; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 2 KVAKOEKKKKTKGRAKR 18
Db 107 QVAKOEKKKKRGRGAKR 123
:IIII:II: I III

RESULT 5	RS30_SCHPO	STANDARD:	PRT:	61 AA.
AC	RS30_SCHPO	042952: 014314.		
DT	15-DEC-1998	(Rel. 37, Created)		
DT	30-MAY-2000	(Rel. 39, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	40S ribosomal protein S30.			
GN	(RS30A OR RS30B OR SPAC19H12.04) OR (RS30B OR SPBC19G7.03C).			
OS	Schizosaccharomyces pombe (Fission yeast).			
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;			
OC	Schizosaccharomyces.			
OX	NCBI_TaxID=4896;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=972;			
RA	Lenaers G., Perret E., Bonnet C., Caput D., Picard A.;			
RT	"Transcription of ribosomal genes is down regulated by ammonium			
RL	starvation in fission yeast."			
RL	Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A. (RS30A AND RS30B).			
RC	STRAIN=972;			
RX	MEDLINE=21848401: Pubmed-11859360;			
RA	Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,			
RA	Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,			
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,			
RA	Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,			
RA	Gentles S., Gobie A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,			
RA	Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,			
RA	James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,			
RA	Mooney P., Moule S., Mungall K., Murphy L., Mblett D., Odell C.,			
RA	Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,			
RA	Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,			
RA	Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,			
RA	Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,			
RA	Woodward J., Voiclaert G., Aert R., Robben J., Grymoprez B.,			
RA	Weltjens I., Vanstreels E., Rieger M., Scheeler M., Moeller-Auer S.,			
RA	Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,			
RA	Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,			
RA	Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,			
RA	Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motlier S.,			
RA	Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,			
RA	Lucas M., Rochet M., Galliardin C., Tallada V.A., Garzon A., Rhode G.,			
RA	Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,			
RA	Domínguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,			
RA	Cerruti L., Lome T., McComble M.R., Paulsen I., Potashkin J.,			
RA	Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;			
RT	"The genome sequence of Schizosaccharomyces pombe."			
RL	Nature 415:871-880(2002)			
CC	-1- MISCELLANEOUS: THERE ARE TWO GENES FOR S30 IN S.POMBE.			
CC	-1- SIMILARITY: BELONGS TO THE S30E FAMILY OF RIBOSOMAL PROTEINS.			
CC	-----			
CC	CC This SWISS-PROT entry is copyright. It is produced through a collaboration -			
CC	CC between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	CC the European Bioinformatics Institute. There are no restrictions on its			
CC	CC use by non-profit institutions as long as its content is in no way			
CC	CC modified and this statement is not removed. Usage by and for commercial			
CC	CC entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	CC or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: AJ002731: CA005693.1: -			
DR	EMBL: AL330814: CAC00552.1: -			
DR	EMBL: AL021839: CAA17057.2: -			
KW	Ribosomal protein; Multigene family.			
SO	SEQUENCE 61 AA; 6910 MW; 4B9C171F1A326F22 CRC64;			
Query Match	53.1%;	Score 51;	DB 1;	Length 61;
Best Local Similarity	61.1%;	Pred. NO. 0.98;		
Matches 11;	Conservative 2;	Mismatches 5;	Indels 0;	Gaps 0;

ID	RS30_ORYLA	STANDARD:	PRT:	59 AA.
AC	09W6YO:			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	40S ribosomal protein S30.			
GN	FAU OR RPS30.			
OS	Oryzias latipes (Medaka fish) (Japanese ricefish).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;			
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;			
OC	Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.			
OX	NCBI_TaxID=8090;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20070552; PubMed=10603084;			
RA	Henrich T., Witbrodt J.;			
RT	"An in situ hybridization screen for the rapid isolation of			
RT	differentially expressed genes.";			
RL	Dev. Genes Evol. 210:28-33(2000).			
CC	-I- MISCELLANEOUS: THIS RIBOSOMAL PROTEIN IS SYNTHESIZED AS A			
CC	C-TERMINAL EXTENSION PROTEIN (CEP) OF A UBQUITIN-LIKE PROTEIN.			
CC	-I- SIMILARITY: BELONGS TO THE S30E FAMILY OF RIBOSOMAL PROTEINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL; AJ238274; CAB40969.1; ALT_INIT.			
DR	Ribosomal protein.			
RW	SEQUENCE 59 AA: 6660 MW; E91B9C088C7772CF CRC64;			
SQ				
Query Match		52.1%;	Score 50;	DB 1; Length 59;
Best Local Similarity		58.8%;	Pred. NO. 1.3;	
Matches 10: Conservative		4;	Mismatches 3;	Indels 0; Gaps 0;
OY	3 VAKOEKKKKKTGRARR 19			
DB	19 VDKHEKEEDGRARR 35			
RESULT 7				
SN24_HUMAN				
AC	P51532:	STANDARD:	PRT:	1647 AA.
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Possible global transcription activator SNF2L4 (SNF2-beta) (BRG-1			
DE	protein) (Mitotic growth and transcription activator) (Brahma			
DE	protein homolog 1).			
GN	SNARCA4 OR SNF2L4 OR BRG1 OR SNF2B.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94050144; PubMed=8232556;			
RA	Khavari P.A., Peterson C.L., Tamkun J.W., Mendel D.B., Crabtree G.R.;			
RT	"BRG1 contains a conserved domain of the SWI2/SNF2 family necessary			
RT	for normal mitotic growth and transcription.";			


```

RL Nature 366:170-174(1993).
RN [2]
RP REVISIONS.
RA Khavari P.A., Peterson C.L., Tamkun J.W., Mendel D.B., Crabtree G.R.;
RN Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RX [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=94268902; PubMed=8208605;
RA Chiba H., Muramatsu M., Nomoto A., Kato H.;
RT "Two human homologues of Saccharomyces cerevisiae SWI2/SNF2 and
RT Drosophila brachma are transcriptional coactivators cooperating with
RT the estrogen receptor and the retinoic acid receptor.";
RL Nucleic Acids Res. 22:1815-1820(1994).
RN [4]
RP SEQUENCE OF 814-1474 FROM N.A.
RA Lamerdin J.E., McCreedy P.M., Showronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Gordon L., Dias J., Brower A., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
RA Dangnanan L., Erlar A., Christensen M., Georgescu A., Avila J., Liu S.,
RA Attix C., Andreise T., Frankheim M., Amico-Keller G., Coeffield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Carraro A.V.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: TRANSCRIPTIONAL COACTIVATOR COOPERATING WITH NUCLEAR
CC HORMONE RECEPTORS TO POTENTIATE TRANSCRIPTIONAL ACTIVATION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U29175; AAB40977.1; -.
DR EMBL: D26156; BAA05143.1; -.
DR EMBL: AC006127; AAC97987.1; -.
DR Genew: HGNC:11100; SMARCA4.
DR MIM: 603254; -.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR000330; SNF2_N.
DR Pfam: PF00176; SNF2_N; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00439; bromodomain; 1.
DR PRINTS: PRO0503; BROMODOMAIN.
DR SMART: SM00297; BROMO; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELICC; 1.
DR PROSITE: PS00633; BROMODOMAIN_1; 1.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
KW Transcription regulation; Nuclear protein; Activator; Bromodomain;
KW ATP-binding; Helicase
FT DOMAIN 578 588 POLY-LYS.
FT DOMAIN 663 672 POLY-GLU.
FT NP_BIND 779 786 ATP (POTENTIAL).
FT SITE 881 884 DEGH BOX.
FT DOMAIN 1360 1364 POLY-GLU.
FT DOMAIN 1477 1547 BROMODOMAIN.
FT DOMAIN 1571 1584 POLY-GLU.
SQ SEQUENCE 1647 AA: 184585 MW: 78785E7953277F1D CRC64;
Query Match 52.1%; Score 50; DB 1; Length 1647;
Best Local Similarity 58.8%; Pred. No. 25;
Matches 10; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
OY 1 AKVAKQEKKKKKTKRAK 17
1:||||:||||:|:

```

```

Db 574 AQVAKKKKKKKKAE 590
RESULT 8
MS22_DROHY STANDARD; PRT: 1391 AA.
ID MS22_DROHY
AC 008696;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Axoneme-associated protein ms101(2).
GN MS101(2).
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7224;
[1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=95045538; PubMed=7957199;
RA Neesen J., Padmanabhan S., Benemann H.;
RT "Tandemly arranged repeats of a novel highly charged 16-amino-acid
RT motif representing the major component of the sperm-tail-specific
RT axoneme-associated protein family Dms101 form extended
RT alpha-helical rods within the extremely elongated spermatozoa of
RT Drosophila hydei.";
RL Eur. J. Biochem. 225:1089-1095(1994).
CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCYTES AND EARLY
CC SPERMATIDS.
CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
CC -1- POLYMORPHISM: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT
CC STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM
CC REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X73481; CAA51876.1; -.
DR PIR: S34154; S34154.
DR FlyBase: FBgn0020733; Dhydymst101(2).
DR Sperm: Repeat; Multigene family; Polymorphism.
FT FT 59 X 16 AA APPROXIMATE TANDEM REPEATS OF
FT DOMAIN 332 1268 [KR]-K-X-C-X-X-A-K-X-K-X-X-X-E.
SQ SEQUENCE 1391 AA: 159000 MW: 1B2A368F30F48B78 CRC64;
Query Match 51.6%; Score 49.5; DB 1; Length 1391;
Best Local Similarity 44.0%; Pred. No. 25;
Matches 11; Conservative 5; Mismatches 2; Indels 7; Gaps 1;
OY 2 KVAKQEKK-----KKTKRAKRR 19
1:||||| 11|:|:|:
Db 943 KLAKEKKAGEKKKLKRAKGKK 967
RESULT 9
RS30_YEAST STANDARD; PRT: 62 AA.
ID RS30_YEAST
AC Q12087;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 40S ribosomal protein S30.
GN (RS30A OR YLR287BC OR LB003.23) AND (RS30B OR YOR192C).
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

```

```

OX  NCBI_TaxID=4932;
RN  [1]
RA  SEQUENCE FROM N.A., AND SEQUENCE (RPS30A).
RC  STRAIN=YRB141;
RX  MEDLINE=96278780; PubMed=8662789;
RA  Baker R.T., Williamson N.A., Wettenhall R.E.H.;
RT  "The yeast homolog of mammalian ribosomal protein S30 is expressed
RT  from a duplicated gene without a ubiquitin-like protein fusion
RT  sequence. Evolutionary implications.";
RL  J. Biol. Chem. 271:13549-13555(1996).
RN  [2]
RA  SEQUENCE FROM N.A. (RPS30A).
RA  Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA  Favell A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
RA  Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
RA  Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,
RA  Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,
RA  Taich A., Trevasakis E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,
RA  Wilson R., Waterston R.;
RL  Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
RN  [3]
RA  SEQUENCE FROM N.A. (RPS30B).
RA  Hughes B., Pohl T.M.;
RL  Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.
CC  -I- MSSS SPECTROMETRY: MW=6987; MW_ERR=3.4; METHOD=MALDI.
CC  -I- SIMILARITY: BELONGS TO THE S30E FAMILY OF RIBOSOMAL PROTEINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@sib-sib.ch).
CC  -----
DR  EMBL; U48700; AAC49317.1; -
DR  EMBL; U48699; AAC49316.1; -
DR  EMBL; U83406; AAB41050.1; -
DR  EMBL; U83407; AAB41051.1; -
DR  EMBL; U17243; AAB67333.1; -
DR  EMBL; Z75090; CAA93931.1; -
DR  SGD; S0004278; RPS30A.
DR  SGD; S0005708; RPS30B.
KW  Ribosomal protein; Multigene family.
FT  INIT_MET 0
FT  SEQUENCE 62 AA: 6987 MW: 65EADADBD8DFC586 CRC64;
SQ

```

```

RC  STRAIN=SL278B;
RA  Smart W.C., Park H.-D., Cooper T.G.;
RL  Submitted (XXX-1994) to the EMBL/Genbank/DBJ databases.
RN  [2]
RA  SEQUENCE FROM N.A.
RX  MEDLINE=95011581; PubMed=7926768;
RA  Strich R., Surosky R.T., Steber C., Messenguy F., Dubois E.,
RA  Easton Espósito R.;
RT  "UME6 is a key regulator of nitrogen repression and meiotic
RT  development.";
RL  Genes Dev. 8:796-810(1994).
RN  [3]
RA  SEQUENCE FROM N.A.
RA  STRAIN=GRE88;
RA  Kuno A.;
RL  Submitted (DEC-1993) to the EMBL/Genbank/DBJ databases.
RN  [4]
RA  SEQUENCE FROM N.A.
RA  STRAIN=S288C / AB972;
RC  Oliver K., Shore L., Harris D., Barrell B.G., Rajandream M.A.,
RA  Walsh S.V.;
RL  Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.
RN  [5]
RA  CHARACTERIZATION.
RX  MEDLINE=92253412; PubMed=1579492;
RA  Park H.-D., Luche R.M., Cooper T.G.;
RT  "The yeast UME6 gene product is required for transcriptional
RT  repression mediated by the CAR1 URS1 repressor binding site.";
RL  Nucleic Acids Res. 20:1909-1915(1992).
RN  [6]
RA  CHARACTERIZATION.
RX  MEDLINE=96100650; PubMed=8528081;
RA  Anderson S.F., Steber C.M., Easton Espósito R., Coleman J.E.;
RT  "UME6, a negative regulator of meiosis in Saccharomyces cerevisiae,
RT  contains a C-terminal Zn2Cys6 binuclear cluster that binds the URS1
RT  DNA sequence in a zinc-dependent manner.";
RL  Protein Sci. 4:1832-1843(1995).
RN  [7]
RA  CHARACTERIZATION.
RX  MEDLINE=96188868; PubMed=8614637;
RA  Jackson J.C., Lopes J.M.;
RT  "The yeast UME6 gene is required for both negative and positive
RT  transcriptional regulation of phospholipid biosynthetic gene
RT  expression.";
RL  Nucleic Acids Res. 24:1322-1329(1996).
CC  -I- FUNCTION: BINDS TO THE URS1 SITE (5'-AGCGCCGCA-3') AND NEGATIVELY
CC  REGULATES THE EXPRESSION OF MANY GENES INCLUDING CAR1 (ARGINASE),
CC  SEVERAL REQUIRED FOR SPOULATION, MATING TYPE SWITCHING, INOSITOL
CC  METABOLISM, AND OXIDATIVE CARBON METABOLISM. HAS BOTH A POSITIVE
CC  AND NEGATIVE ROLE IN REGULATING PHOSPHOLIPID BIOSYNTHESIS.
CC  -I- SUBUNIT: INTERACTS WITH IME1.
CC  -I- SUBCELLULAR LOCATION: Nuclear.
CC  -I- SIMILARITY: CONTAINS 1 ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
CC  CLUSTER DOMAIN.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@sib-sib.ch).
CC  -----
DR  EMBL; L32186; AAA34471.1; -
DR  EMBL; L24539; AAC14472.1; -
DR  EMBL; D23663; BAA04890.1; -
DR  EMBL; Z68194; CAA92346.1; -
DR  HSSP; P07272; IPLY1.
DR  TRNSPAC; T01247; -
DR  SGD; S0002615; UME6.
DR  InterPro: IPR001138; Fungi_Trn.
DR  Pfam; PF00172; Zn_c1us; 1.
DR  SMART; SM00066; GAL4; 1.

```

```
DR PROSITE; PS00463; ZN2_CY6_FUNGAL_1; 1.
DR PROSITE; PS50048; ZN2_CY6_FUNGAL_2; 1.
KM Transcription regulation: Repressor: Activator: DNA-binding:
KW Nuclear protein; Zinc; Metal-binding.
FT DNA_BIND 771 798 ZN(2)-CYS(6), FUNGAL-TYPE.
FT CONFLICT 101 101 V -> G (IN REF. 3).
FT CONFLICT 363 363 I -> V (IN REF. 1).
FT CONFLICT 443 443 N -> T (IN REF. 1).
FT CONFLICT 465 465 G -> D (IN REF. 1).
SQ SEQUENCE 836 AA; 91123 MW; 0DDA0A6B4A157182 CRC64;

Query Match 51.0%; Score 49; DB 1; Length 836;
Best Local Similarity 63.2%; Pred. NO. 19;
Matches 12; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 KVAQEKKKKKTKGRARRA 20
| 1: 1111 1111
DB 814 KMKLEIKTKTKKRAARRA 832

RESULT 11
H14_RABIT STANDARD; PRT; 73 AA.
ID H14_RABIT
AC P02252;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1.4 (Fragment).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OC NCBI_TaxID=9986;
RN 11
RP SEQUENCE.
RX MEDLINE=72068710; PubMed=5167020;
RA Rall S.C., Cole R.D.;
RT "Amino acid sequence and sequence variability of the amino-terminal
RT regions of lysine-rich histones."
RL J. Biol. Chem. 246:7175-7190(1971).
RN 12
RP SEQUENCE.
RX MEDLINE=71134818; PubMed=5547708;
RA Langen T.A., Rall S.C., Cole R.D.;
RT "Variation in primary structure at a phosphorylation site in
RT lysine-rich histones."
RL J. Biol. Chem. 246:1942-1944(1971).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
DR PIR: A02579; A02579.
DR HSSP: P08287; IGHC.
DR InterPro: IPR001386; Histone_H1/H5.
DR Pfam: PF00538; linker_histone_1.
DR SMART: SM00526; H1S; 1.
KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW Acetylation; Phosphorylation.
FT MOD_RES 1 1 ACETYLATION.
FT MOD_RES 37 37 PHOSPHORYLATION.
FT DOMAIN 37 >73 GLOBULAR.
FT NON_TER 73 73
SQ SEQUENCE 73 AA; 7182 MW; 03875E24BE281F94 CRC64;

Query Match 50.0%; Score 48; DB 1; Length 73;
Best Local Similarity 50.0%; Pred. NO. 2.9;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 AKVAKKKKKTKGRARRA 20
| 1: 111 111:1
DB 17 ATPVKAKKKKSAGAKRKA 36

RESULT 12
```

```
YBX6_SCHPO STANDARD; PRT; 578 AA.
ID YBX6_SCHPO
AC 010202;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative ATP-dependent RNA helicase C17D1.06.
DE SPC17D1.06.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetes; Schizosaccharomycetaceae;
OC NCBI_TaxID=4896;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tvey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voicakeart G., Aert R., Robben J., Grymonprez B.,
RA Weltjens E., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motter S.,
RA Gallibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AL031322; CAA20430.1; -.
CC HSSP: Q58083; IHV8.
CC DR InterPro: IPR001410; DEAD.
CC DR InterPro: IPR000629; DEAD_box.
CC DR InterPro: IPR001650; Helicase_C.
CC DR Pfam: PF00270; DEAD_1.
CC DR Pfam: PF00271; helicase_C; 1.
CC DR SMART: SM00487; DEXDC; 1.
CC DR SMART: SM00490; HELIC_C; 1.
CC DR PROSITE: PS00039; DEAD_ATP_HELICASE; 1.
KW Hypothetical protein; ATP-binding; RNA-binding; Helicase.
FT NP_BIND 209 216 ATP (POTENTIAL).
FT SITE 316 319 DEAD BOX.
SQ SEQUENCE 578 AA; 63912 MW; 0A84EC77ECC29FD CRC64;

Query Match 50.0%; Score 48; DB 1; Length 578;
Best Local Similarity 38.9%; Pred. NO. 18;
Matches 7; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 2 KVAQEKKKKKTKGRARR 19
```

Db 93 RISKDKKKKKKKSKTK 110

RESULT 13

VG01_HSV11 STANDARD: PRT: 862 AA.

AC 000132:

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 01-DEC-1992 (Rel. 24, Last annotation update)

DE Hypothetical gene 1 protein.

GN 1.

OS Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Ictalurid Herpes-like viruses.

OX NCBI_TaxID=10401;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Auburn 1;

RX MEDLINE=92087490; PubMed=1727613;

RA Davison A.J.;

RT "Channel catfish virus: a new type of herpesvirus.";

RU Virology 186:9-14(1992).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: M75136; AAA8182.1;

DR EMBL: M75136; AAA8104.1;

DR PIR: B36786; B36786.

KW Hypothetical protein.

SO SEQUENCE 862 AA; 93511 MW; 087E153EECEC71AD6 CRC64;

OY 1 AKVAKOEKKKKKTGRK 17

Db 691 AKVTKEKTEKTPQTK 707

RESULT 14

IF2P_YEAST

AC P39730: STANDARD: PRT: 1002 AA.

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Translation Initiation Factor IF-2.

GN YAL035W OR FUN12.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C / AB972;

RX MEDLINE=95249563; PubMed=7731988;

RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,

RA Hall U., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,

RA Storms R.K.;

RT "The nucleotide sequence of chromosome I from Saccharomyces

RT cerevisiae.";

RT Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=94357438; PubMed=8076820;

RA Suttrave P., Shafer B.K., Strathern J.N., Hughes S.H.;

RT "Isolation, identification and characterization of the FUN12 gene of

RT Saccharomyces cerevisiae.";

RL Gene 146:209-213(1994).

RN [3]

RP REVISIONS.

RC STRAIN=S288C / AB972;

RA Vo D.;

RL Submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.

RN [4]

RP CHARACTERIZATION.

RX MEDLINE=98288357; PubMed=9624054;

RA Choi S.K., Lee J.H., Zoll W.L., Merrick W.C., Dever T.E.;

RT "Promotion of met-trNAI Met binding to ribosomes by yif2, a bacterial

RT IF2 homolog in yeast.";

RL Science 280:1757-1760(1998).

CC -I- FUNCTION: FUNCTION IN GENERAL TRANSLATION INITIATION BY PROMOTING

CC THE BINDING OF THE FORMYL METHIONINE-tRNA TO RIBOSOMES. SEEMS TO

CC -I- SIMILARITY: BELONGS TO THE IF-2 FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U12980; AAC04996.1;

DR EMBL: L29389; AAA57228.1; ALT_SEQ.

DR SGD: S0000033; FUN12.

DR InterPro: IPR004161; EFTU_D2.

DR InterPro: IPR000795; EF_GTPbind.

DR InterPro: IPR000178; IF2.

DR InterPro: IPR005225; Small_GTP.

DR Pfam: PF03144; GTP_EFTU_D2; 2.

DR Pfam: PF03144; GTP_EFTU_D2; 2.

DR TIGRfams: TIGR00231; small_GTP; 1.

DR PROSITE: PS01176; IF2; FALSE NEG.

KW Initiation factor; Protein biosynthesis; GTP-binding.

FT DOMAIN 361 371

FT NP_BIND 412 419

FT POLY_GLU.

SO SEQUENCE 1002 AA; 112268 MW; 1A496195DAE1C283 CRC64;

OY 1 AKVAKOEKKKKKTGRARRA 20

Db 215 AKAKKEKKEKAK--REKRA 232

RESULT 15

STP2_RAT

AC P11101: STANDARD: PRT: 116 AA.

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 01-FEB-1996 (Rel. 33, Last annotation update)

DE Nuclear transition protein 2 (TP-2).

GN TNP2.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=89263797; PubMed=2726489;

RA Luerksen H., Maier W.M., Hoyer-Fender S., Engel W.;

RT "The nucleotide sequence of rat transition protein 2 (TP2) cDNA.";

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2003, 17:00:45 ; Search time 15 Seconds
(without alignments)
128.179 Million cell updates/sec

Title: US-09-424-815E-6
Perfect score: 96
Sequence: 1 AKVAKOEKKKKTKGRAKRRA 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	88	91.7	133	1 A47416	ubiquitin-like pro
2	88	91.7	133	2 JC1278	ubiquitin-like pro
3	88	91.7	133	1 I48346	ribosomal protein
4	65	67.7	130	2 T15642	hypothetical prote
5	64	66.7	62	2 H85342	RIBOSOMAL PROTEIN
6	64	66.7	68	2 F84580	40S ribosomal prot
7	60	62.5	58	2 A71604	ribosomal protein
8	55	57.3	439	2 J01729	ankyrin-repeat pro
9	51	53.1	61	2 T39834	ribosomal protein
10	50	52.1	1051	2 JC4091	glycoprotein A - P
11	50	52.1	1647	2 S45252	SNF2beta protein -
12	49.5	51.6	1390	2 S51364	sperm tail-specific
13	49	51.0	63	1 S67074	ribosomal protein
14	49	51.0	85	2 T26143	hypothetical prote
15	49	51.0	647	2 T23407	hypothetical prote
16	49	51.0	836	2 S61570	regulatory protein
17	48	50.0	73	2 A02579	histone H1.4 - rab
18	48	50.0	207	2 D66225	hypothetical prote
19	48	50.0	578	2 S67386	probable ATP-depen
20	48	50.0	862	2 B36786	hypothetical prote
21	48	50.0	1002	2 S70292	FUN12 protein - ye
22	47.5	49.5	894	2 T15769	hypothetical prote
23	47	49.0	113	2 S57667	translational prote
24	47	49.0	183	2 F84636	hypothetical prote
25	47	49.0	439	2 E84780	hypothetical prote
26	47	49.0	519	2 S14529	translational prote
27	47	49.0	645	1 T37740	serine proteinase
28	47	49.0	649	1 T37740	coiled-coil protei
29	47	49.0	785	2 B84038	DNA mismatch repa

30	46.5	48.4	1017	2 T48452	hypothetical prote
31	46	47.9	131	2 G69959	hypothetical prote
32	46	47.9	218	1 HSC81	histone H1.02 - ch
33	46	47.9	218	2 A23055	histone H1.01 - ch
34	46	47.9	218	2 S01262	histone H1 - musco
35	46	47.9	266	2 T52089	ribosomal protein
36	46	47.9	392	2 T15755	hypothetical prote
37	46	47.9	409	2 T24543	hypothetical prote
38	46	47.9	545	2 T08564	hypothetical prote
39	45.5	47.4	277	2 T34625	probable MLP/p60 f
40	45.5	47.4	478	2 T51521	hypothetical prote
41	45	46.9	215	2 I52523	nucleoporin p62 ho
42	45	46.9	305	2 T24129	hypothetical prote
43	45	46.9	316	2 T08694	hypothetical prote
44	45	46.9	441	2 A48455	acidic phosphoprot
45	45	46.9	451	2 G01227	receptin - human

ALIGNMENTS

RESULT 1
A47416
ubiquitin-like protein / ribosomal protein S30, cytosolic (validated) - rat
N:Contains: ribosomal protein S30; ubiquitin-like protein
C:Species: Rattus norvegicus (Norway rat)
C>Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: A47416; B47416; S18101
J. Olivera, J.: Wool, I.G.
A:Title: The carboxyl extension of a ubiquitin-like protein is rat ribosomal protein
A:Reference number: A47416; MUID:93352612; PMID:8394356
A:Accession: A47416
A:Molecule type: mRNA
A:Residues: 1-133 <OLV>
A:Cross-references: EMBL:X62671; NID:9407165; PIDN:CAA44545.1; PID:957566
A:Accession: B47416
A:Molecule type: Protein
A:Residues: 75-92 <OL2>
A>Note: The proteins are designated as ubiquitin-like protein and ribosomal protein S
C:Superfamily: ubiquitin-like protein / rat ribosomal protein S30; ubiquitin homology
C:Keywords: protein biosynthesis; ribosome
F:1-74/Product: ubiquitin-like protein #status predicted <UBI>
F:1-74/Domains: ubiquitin homology <UBH>
F:75-133/Product: ribosomal protein S30 #status experimental <RIB>
Query Match 91.7%; Score 88; DB 1; Length 133;
Best Local Similarity 100.0%; Pred. No. 0.00013;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 KVAKOEKKKKTKGRARR 19
DB 92 KVAKOEKKKKTKGRARR 109
IIIIIIIIIIIIIIIIIIII
ubiquitin-like protein / ribosomal protein S30, cytosolic - human
N:Alternate names: fau protein
C:Species: Homo sapiens (man)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 24-Sep-1999
C:Accession: JC1278; I37387; S68911; S21449; S21450
R:Kas, K.; Michiels, L.; Merregaert, J.
Biochem. Biophys. Res. Commun. 187, 927-933, 1992
A:Title: Genomic structure and expression of the human fau gene: Encoding the ribosome
A:Reference number: JC1278; MUID:92412144; PMID:1326960
A:Accession: JC1278
A:Molecule type: DNA
A:Residues: 1-133 <KAS>
A:Cross-references: EMBL:X65921; NID:931304; PIDN:CAA46714.1; PID:931305
R:Michiels, L.; Van der Raaij-Majumder, E.; Van Hasselt, F.; Kas, K.; Merregaert, J.
Oncogene 8, 2537-2546, 1993
A:Title: fau cDNA encodes a ubiquitin-like-S30 fusion protein and is expressed as an

A:Reference number: I37387; MUID:93368957; PMID:8395683
A:Accession: I37387
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-133 <RES>
A:Cross-references: EMBL:X65923; NID:g31302; PIDN:CAA46716.1; PID:g31303
R:Vladimirov, S.N.; Ivanov, A.V.; Karpova, G.G.; Musolyanov, A.K.; Egorov, T.A.; Thiede,
Eur. J. Biochem. 239, 144-149, 1996
A>Title: Characterization of the human small-ribosomal-subunit proteins by N-terminal an
A:Reference number: S68911; MUID:96305378; PMID:8706659
A:Accession: S68911
A:Molecule type: protein
A:Residues: 75-99 <VLA>
C:Genetics:
A:Gene: fau
A:Introns: 25/3; 74/1; 92/3
C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology
C:Keywords: protein biosynthesis; ribosome
F:1-74/Domain: ubiquitin homology <DBH>
F:75-133/Product: ribosomal protein S30, cytosolic #status experimental <MAT>

Query Match 91.7%; Score 88; DB 2; Length 133;
Best Local Similarity 100.0%; Pred. No. 0.00013;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KVAKEKKKKKTGRAKR 19
DB 92 KVAKEKKKKKTGRAKR 109
|||||

RESULT 3
I48346
ribosomal protein fau - mouse
N:Alternate names: gene fau protein; monoclonal nonspecific suppressor factor beta
C:Species: Mus musculus (house mouse)
C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 24-Sep-1999
C:Accession: I48346; A56532; I59368; S21452
R:Michiels, L.; Van der Raaij-Waaij, E.; Van Hasselt, F.; Kas, K.; Merregaert, J.
Oncogene 8, 2537-2546, 1993
A>Title: fau cDNA encodes a ubiquitin-like-S30 fusion protein and is expressed as an an
A:Reference number: I37387; MUID:93368957; PMID:8395683
A:Accession: I48346
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-133 <RES>
A:Cross-references: EMBL:X65922; NID:g50949; PIDN:CAA46715.1; PID:g50950
R:Castels, D.; Poitier, C.; Guenet, J.L.; Merregaert, J.
Genomics 25, 291-294, 1995
A>Title: The mouse fau gene: genomic structure, chromosomal localization, and character
A:Reference number: A56532; MUID:95293388; PMID:7774934
A:Accession: A56532
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-133 <CAS>
A:Cross-references: GB:I3715; NID:g497610; PIDN:AA91564.1; PID:g497611
A>Note: authors translated the codon GTP for residue 119 as Arg, and GTC for residue 120
R:Nakamura, M.; Xavier, R.M.; Tsunematsu, T.; Tanigawa, Y.
Proc. Natl. Acad. Sci. U.S.A. 92, 3463-3467, 1995
A>Title: Molecular cloning and characterization of a cDNA encoding monoclonal nonspecific
A:Reference number: I59368; MUID:95241522; PMID:7724584
A:Accession: I59368
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-133 <RE2>
A:Cross-references: GB:ID26610; NID:g1060926; PIDN:BA05655.1; PID:g1060927
C:Genetics:
A:Gene: fau
A:Introns: 25/3; 74/1; 92/3
C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology
F:1-74/Domain: ubiquitin homology <UBH>

Query Match 91.7%; Score 88; DB 2; Length 133;
Best Local Similarity 100.0%; Pred. No. 0.00013;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KVAKEKKKKKTGRAKR 19
DB 92 KVAKEKKKKKTGRAKR 109
|||||

RESULT 4
T15642
hypothetical protein C26F1.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 24-Nov-1999
C:Accession: T15642
R:Geisel, C.; Steiljes, L.; Bradshaw, H.
submitted to the EMBL Data Library, March 1996
A:Description: The sequence of C. elegans cosmid C26F1.
A:Reference number: Z18381
A:Accession: T15642
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-130 <GEI>
A:Cross-references: EMBL:U53148; NID:g1255375; PID:g1255381; PIDN:AA837076.1; GSPDB:G
A:Experimental source: strain Bristol NZ; clone C26F1
C:Genetics:
A:Gene: CESP:C26F1.4
A:Map position: 5
A:Introns: 27/3; 71/1
C:Superfamily: unassigned ubiquitin-related proteins; ubiquitin homology

Query Match 67.7%; Score 65; DB 2; Length 130;
Best Local Similarity 77.8%; Pred. No. 0.11;
Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 KVAKEKKKKKTGRAKR 19
DB 89 KVDKDKKKKRGRAKRR 106
|||||

RESULT 5
H85342
RIBOSOMAL PROTEIN S30 homolog [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: H85342
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold sp
Nature 402, 769-777, 1999
A>Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488; PMID:10617198
A:Accession: H85342
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-62 <STO>
A:Cross-references: GB:NC_001268; NID:g7269837; PIDN:CAB79697.1; GSPDB:GN00140
C:Genetics:
A:Gene: Atg29390
A:Map position: 4

Query Match 66.7%; Score 64; DB 2; Length 62;
Best Local Similarity 72.2%; Pred. No. 0.085;
Matches 13; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 KVAKEKKKKKTGRAKR 19
DB 20 KVAKEKKKKKTGRAKR 37
|||||

RESULT 6
F84580
40S ribosomal protein S30 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 17-May-2002
C:Accession: F84580
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y

M.: Koo, H.; Koffel, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umeyam, L.; Tallon, L.; Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999

A>Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A:Reference number: A844420; MUID:20083487; PMID:10617197

A:Accession: F84580

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-68 <STO>

C:Cross-references: GB:AE002093; NID:g3687243; PIDN:AA62141.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g19750

A:Map position: 2

C:Superfamily: yeast ribosomal protein S30.e

Query Match 66.7%; Score 64; DB 2; Length 68;

Best Local Similarity 72.2%; Pred. No. 0.091;

Matches 13; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Oy 2 KVAKEKKKKKTGRAKR 19

|||||:||||| ||| :|

Db 26 KVAKODKKKKPRGRAKR 43

RESULT 7

A1604 ribosomal protein S30 PFB0885w - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 13-Aug-1999

C:Accession: A71604

R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O. Science 282, 1126-1132, 1998

A>Title: Chromosome 2, sequence of the human malaria parasite *Plasmodium falciparum*.

A:Reference number: A71600; MUID:99021743; PMID:9804551

A:Accession: A71604

A>Status: Preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-58 <GAR>

A:Cross-references: GB:AE001422; GB:AE001362; NID:g3845298; PIDN:AC71966.1; PID:g384530

A:Experimental source: clone 3D7

C:Genetics:

A:Gene: PFB0885w

C:Superfamily: yeast ribosomal protein S30.e

Query Match 62.5%; Score 60; DB 2; Length 58;

Best Local Similarity 66.7%; Pred. No. 0.26;

Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 2 KVAKEKKKKKTGRAKR 19

||| :||||| :|||||

Db 20 KVPKDKKKRKTGRAKR 37

RESULT 8

J01729

ankyrin-repeat protein - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 22-Jun-1999

C:Accession: J01729

R:Zhang, H.; Schejter, D.C.; Fowle, W.H.; Goodman, H.M. Plant Cell 4, 1575-1588, 1992

A>Title: Expression of antisense RNA of an ankyrin repeat-containing gene block

A:Reference number: J01729; MUID:93104681; PMID:1281700

A:Accession: J01729

A:Molecule type: mRNA

A:Residues: 1-439 <ZHA>

A:Cross-references: GB:M82883; NID:g166743; PIDN:AA32812.1; PID:g166744

C:Genetics:

A:Gene: AKR

A:Introns: 262/1; 287/3; 320/3; 353/3; 386/3; 419/3

C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology

C:Keywords: cytoskeleton

F:282-314/Domains: ankyrin repeat <RP1>

F:315-347/Domains: ankyrin repeat <RP2>

F:348-380/Domains: ankyrin repeat <RP3>

F:381-413/Domains: ankyrin repeat <RP4>

F:387-419/Domains: ankyrin repeat homology <AN1>

Query Match 57.3%; Score 55; DB 2; Length 439;

Best Local Similarity 64.7%; Pred. No. 6;

Matches 11; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 2 KVAKEKKKKKTGRAKR 18

:|||||:| :| |||

Db 107 QVAKEKKKKRRGCAKR 123

RESULT 9

T39834 ribosomal protein s30 - fission yeast (*Schizosaccharomyces pombe*)

C:Species: *Schizosaccharomyces pombe*

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Jun-2000

C:Accession: T39834

R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Devlin, K.; Churcher, C.M. submitted to the EMBL Data Library, February 1998

A:Reference number: 221884

A:Accession: T39834

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-61 <LYN>

A:Cross-references: EMBL:AL021839; PIDN:CA17057.2; GSPDB:GN00067; SPDB:SPBC1967.03c

A:Experimental source: strain 972h-; cosmid c1967

C:Genetics:

A:Gene: SPDB:SPBC1967.03c

A:Map position: 2

A:Introns: 13/3; 56/3

C:Superfamily: yeast ribosomal protein S30.e

Query Match 53.1%; Score 51; DB 2; Length 61;

Best Local Similarity 61.1%; Pred. No. 4;

Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Oy 2 KVAKEKKKKKTGRAKR 19

||| ||||| : ||| :|

Db 20 KVEKEKKPKQPKGRAKR 37

RESULT 10

JC4091

glycoprotein A - *Pneumocystis carinii*

N:Alternate names: gp120; MSG; P15

C:Species: *Pneumocystis carinii*

C:Date: 23-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 15-Jun-2001

C:Accession: JC4091

R:Wright, T.W.; Blisondial, T.V.; Haidaris, C.G.; Gigliotti, F.; Simpson Haidaris, P. DNA Res. 2, 77-88, 1995

A>Title: Isoform diversity and tandem duplication of the glycoprotein A gene in *ferre*

A:Reference number: JC4091; MUID:96093928; PMID:7584051

A:Accession: JC4091

A:Molecule type: mRNA

A:Residues: 1-1051 <WRI>

A:Cross-references: GB:U19871; NID:9841241; PIDN:AA67766.1; PID:9841242

C:Comments: This protein is one of the ligands participating in the adherence of the b

C:Genetics:

A:Gene: gpa

C:Superfamily: *Pneumocystis carinii* major surface glycoprotein MSG100

C:Keywords: glycoprotein; tandem repeat

F:135,259,417,420,514,612,631,642,672,689,703,757,830/Binding site: carbohydrate (Asn

Query Match 52.1%; Score 50; DB 2; Length 1051;

Best Local Similarity 50.0%; Pred. No. 54;

Matches 9; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ANVAKEKKKKKTGRAKR 18

:| :| :| :| :| :|

Db 762 SKLCKKKKKKKKKKK 779

RESULT 11

S45252
SNF2beta protein - human
C:Species: Homo sapiens (man)
C:Date: 10-Dec-1994 #sequence_rev1sion 17-Nov-1995 #text_change 02-Aug-2002
C:Accession: S45252
R:Chiba, H.; Muramatsu, M.; Nomoto, A.; Kato, H.
Nucleic Acids Res. 22, 1815-1820, 1994
A:Title: Two human homologues of Saccharomyces cerevisiae SWI2/SNF2 and Drosophila brachypter
A:Reference number: S45251; MUID:94268902; PMID:8208605
A:Accession: S45252
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1647 <CH1>
A:Cross-references: GB:D26156; NID:9505087; PIDN:BA05143.1; PID:9505088
C:Superfamily: human SNF2alpha protein; bromodomain homology
F:1485-1540/Domain: bromodomain homology <BRO>

Query Match

Best Local Similarity 52.1%; Score 50; DB 2; Length 1647;
Matches 10; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 KVAKQEKKKKKTKGRK 17

Db 574 AQVAKKKKKKKKKKAE 590

RESULT 12

S51364
specm tail-specific protein mst101(2) - fruit fly (Drosophila hydei)
C:Species: Drosophila hydei
C:Date: 19-Jul-1996 #sequence_rev1sion 26-Jul-1996 #text_change 17-Mar-1999
C:Accession: S51364; S34154
R:Neesen, J.; Padmanabhan, S.; Buenemann, H.
Eur. J. Biochem. 225, 1089-1095, 1994
A:Title: Tandemly arranged repeats of a novel highly charged 16-amino-acid motif represen
Alpha-helical rods within the extremely elongated spermatozoa of Drosophila hydei.
A:Reference number: S51364; MUID:95045538; PMID:7957199
A:Accession: S51364
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1390 <NEE>
A:Cross-references: EMBL:X73481
R:Neesen, J.; Heinlein, U.A.O.; Buenemann, H.
submitted to the EMBL Data Library, June 1993
A:Reference number: S34153
A:Accession: S34154
A:Molecule type: DNA
A:Residues: 1-163, 'E', 164-236, 'Q', 237-254, 257-320, 'E', 321-1390 <NEW>
A:Cross-references: EMBL:X73481; NID:g313201; PID:g313202
C:Genetics: mst101(2)
A:Gene: mst101(2)
A:Cross-references: FlyBase:FBgn0011816

Query Match

Best Local Similarity 51.6%; Score 49.5; DB 2; Length 1390;
Matches 11; Conservative 5; Mismatches 2; Indels 7; Gaps 1;

QY 2 KVAQEKKKKKTKGRK 19

Db 942 KLAKEKKKAGEKNLKKKAGKKK 966

RESULT 13

S67074
ribosomal protein S30.e, cytosolic - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein O4725; protein YLR287c-a; protein YOR182c
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence_rev1sion 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: S67074; S70775; S70776; S70777

R:Hughes, B.; Pohl, T.M.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66685
A:Accession: S67074
A:Molecule type: DNA
A:Residues: 1-63 <HUG>
A:Cross-references: EMBL:Z75090; GSPDB:GN00015; MIPS:YOR182c; NID:g1420438; PIDN:CAA9
A:Experimental source: strain S288C
A:Genetics: YS30B
R:Baker, R.T.; Williamson, N.A.; Wettenhall, R.E.H.
J. Biol. Chem. 271, 13549-13555, 1996
A:Title: The yeast homolog of mammalian ribosomal protein S30 is expressed from a dup
A:Reference number: S70774; MUID:96278780; PMID:8662789
A:Accession: S70775
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-63 <BAK>
A:Cross-references: EMBL:U48700; NID:g1256752; PIDN:AAC49317.1; PID:g1256753
A:Genetics: YS30A
A:Accession: S70776
A:Molecule type: mRNA
A:Residues: 1-63 <BAW>
A:Cross-references: EMBL:U48699; NID:g1256750; PIDN:AAC49316.1; PID:g1256751
A:Genetics: YS30A
A:Accession: S70774
A:Molecule type: protein
A:Residues: 2-63 <BA3>
C:Genetics: <YS30B>
A:Gene: SGD:RPS30B; MIPS:YOR182c
A:Cross-references: MIPS:YOR182c; SGD:S0005708
A:Map position: 15R
A:Introns: 1/3
C:Genetics: <YS30A>
A:Gene: SGD:RPS30A; MIPS:YLR287c-a
A:Cross-references: MIPS:YLR287c-a; SGD:S0004278
A:Map position: 12R
A:Introns: 1/3
C:Superfamily: yeast ribosomal protein S30.e
C:Keywords: cytosol; protein biosynthesis; ribosome
F:2-63/Product: ribosomal protein S30.e, cytosolic #status experimental <MAT>

Query Match

Best Local Similarity 61.1%; Score 49; DB 1; Length 63;
Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 KVAQEKKKKKTKGRK 19

Db 20 KVEKTEKKKKPKGRAYKR 37

RESULT 14

T26143
hypothetical protein W04D12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_rev1sion 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26143
R:Miller, N.
submitted to the EMBL Data Library, June 1994
A:Description: The sequence of C. elegans cosmid W04D12.
A:Reference number: Z20160
A:Accession: T26143
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-85 <MIL>
A:Cross-references: EMBL:U00068; PIDN:AAA50745.1; CESP:W04D12.1
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:W04D12.1

Query Match

Best Local Similarity 51.0%; Score 49; DB 2; Length 85;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 KVAQEKKKKKTKGRK 19

Db 20 KVEKTEKKKKPKGRAYKR 37

Db 54 KVAKKKKKKKMKDEKKK 71

RESULT 15

T23407
hypochelical protein K07C5.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T23407
R:McMurray, A.
Submitted to the EMBL Data Library, April 1996
A:Reference number: Z19737
A:Accession: T23407
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-647 <MIL>
A:Cross-references: EMBL:Z71181; PIDN:CAA94899.1; GSPDB:GN00023; CESP:K07C5.6
A:Experimental source: clone K07C5
C:Genetics:
A:Gene: CESP:K07C5.6
A:Map position: 5
A:introns: 36/3; 138/2; 598/3

Query Match 51.0%; Score 49; DB 2; Length 647;
Best Local Similarity 55.6%; Pred. NO. 49;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 2 KVAKQEKKKKTGRAKRR 19
| | : | | | | : | |
Db 537 KREKREKKKAKLGKRRRR 554

Search completed: February 11, 2003, 17:05:21
Job time : 26 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2003, 17:02:55 : Search time 27 Seconds
(without alignments)
18.925 Million cell updates/sec

Title: US-09-424-815E-6
Perfect score: 96
Sequence: 1 AKVAKQKKKKKTGRARRA 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 140259 seqs, 2554876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	91.7	170	10	US-09-925-301-994
2	66	68.8	118	10	US-09-864-761-35333
3	53	55.2	66	10	US-09-764-869-722
4	53	55.2	66	10	US-09-764-846-163
5	51	53.1	59	10	US-09-764-847-823
6	49	51.0	40	10	US-09-925-301-1677
7	49	51.0	145	10	US-09-925-302-572
8	49	51.0	154	10	US-09-925-299-1087
9	48.5	50.5	761	9	US-10-011-582-10
10	48.5	50.5	790	9	US-10-011-582-6
11	48	50.0	62	10	US-09-864-761-39484
12	48	50.0	64	10	US-09-864-761-45501
13	47	49.0	80	10	US-09-864-761-35659
14	47	49.0	521	10	US-09-925-300-1667
15	46.5	48.4	101	9	US-09-738-626-4468
16	46	47.9	66	10	US-09-864-761-39983
17	46	47.9	86	10	US-09-864-761-33832
18	46	47.9	98	10	US-09-764-846-229
19	46	47.9	99	10	US-09-864-761-38013

20	46	47.9	122	10	US-09-925-297-721	Sequence 721, App
21	46	47.9	315	10	US-09-764-853-793	Sequence 793, App
22	45	46.9	46	9	US-09-774-639-261	Sequence 261, App
23	45	46.9	48	10	US-09-764-846-156	Sequence 156, App
24	45	46.9	52	10	US-09-864-761-16109	Sequence 46109, A
25	45	46.9	58	10	US-09-764-846-261	Sequence 261, App
26	45	46.9	63	10	US-09-764-846-185	Sequence 185, App
27	45	46.9	67	12	US-10-001-843-156	Sequence 156, App
28	45	46.9	71	10	US-09-864-761-37061	Sequence 37061, A
29	45	46.9	72	10	US-09-864-761-36199	Sequence 36199, A
30	45	46.9	74	10	US-09-764-846-190	Sequence 190, App
31	45	46.9	74	10	US-09-764-846-262	Sequence 262, App
32	45	46.9	76	9	US-10-002-344A-229	Sequence 229, App
33	45	46.9	76	10	US-09-764-869-960	Sequence 960, App
34	45	46.9	76	10	US-09-764-846-257	Sequence 257, App
35	45	46.9	166	10	US-09-925-299-1088	Sequence 1088, App
36	45	46.9	181	10	US-09-764-846-264	Sequence 264, App
37	45	46.9	311	10	US-09-893-519A-54	Sequence 54, App
38	45	46.9	829	10	US-09-815-242-11216	Sequence 11216, A
39	44	45.8	29	10	US-09-925-301-1616	Sequence 1616, App
40	44	45.8	43	10	US-09-925-301-1633	Sequence 1633, App
41	44	45.8	47	10	US-09-925-297-720	Sequence 720, App
42	44	45.8	47	10	US-09-764-846-952	Sequence 952, App
43	44	45.8	66	10	US-09-764-846-253	Sequence 253, App
44	44	45.8	69	10	US-09-764-869-1222	Sequence 1222, App
45	44	45.8	72	10	US-09-925-300-1105	Sequence 1105, App

ALIGNMENTS

RESULT 1
US-09-925-301-994
; Sequence 994, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925, 301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124, 270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 994
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-994
QY 2 KVAKQKKKKKTGRARR 19
Db 129 KVAKQKKKKKTGRARR 146
RESULT 2
US-09-864-761-35333
; Sequence 35333, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.

Query Match 91.7%: Score 88; DB 10; Length 170;
Best local Similarity 100.0%: Pred. No. 4.5e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aemlica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 35333
LENGTH: 118
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC009704.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 19
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.8
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 18
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 9.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 81
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 12
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 30
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12
OTHER INFORMATION: SWISSPROT HIT: P35544, EVALUE 4.00e-23
OTHER INFORMATION: EST_HUMAN HIT: BR347807.1, EVALUE 6.00e-47
US-09-864-761-35333

```

```

Query Match          68.8%; Score 66; DB 10; Length 118;
Best Local Similarity 72.7%; Pred. No. 0.024;
Matches 16; Conservative 1; Mismatches 1; Indels 4; Gaps 1;

```

```

QY 2 KVAKOEKK---KKKTGRARR 19
   | | | | | | | | | | | | | | | | | |
DB 73 KVAKOEKKKKKKKKKTGQATRR 94

```

```

RESULT 3
US-09-764-869-722
Sequence 722, Application US/09764869
Patent No. US20020061521A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007
CURRENT APPLICATION NUMBER: US/09/764,869
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 2442
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 722
LENGTH: 66
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (4)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (56)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-869-722

```

```

Query Match          55.2%; Score 53; DB 10; Length 66;
Best Local Similarity 55.6%; Pred. No. 0.67;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 2 KVAKOEKKKKKTGRARR 19
   | | | | | | | | | | | | | | | | | |
DB 45 KKKKKKKKKKKKKKKKKRR 62

```

```

RESULT 4
US-09-764-846-163
Sequence 163, Application US/09764846
Patent No. US20020102638A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT212
CURRENT APPLICATION NUMBER: US/09/764,846
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 348
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 163
LENGTH: 66
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (4)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (56)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-846-163

```

```

Query Match          55.2%; Score 53; DB 10; Length 66;
Best Local Similarity 55.6%; Pred. No. 0.67;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 2 KVAKOEKKKKKTGRARR 19
   | | | | | | | | | | | | | | | | | |
DB 45 KKKKKKKKKKKKKKKKKRR 62

```

```

RESULT 5
US-09-764-847-823

```

```
; Sequence 823, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; PRIORITY APPLICATION data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 823
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-847-823

Query Match          53.1%; Score 51; DB 10; Length 59;
Best Local Similarity 55.6%; Pred. No. 1.1;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Oy      2 KVAKOEKKKKKTGRAKR 19
        1-:|||||11-:1
Db      14 KYQKKKKKKKGGRSRSR 31

RESULT 6
US-09-925-301-1677
; Sequence 1677, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1677
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (2)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (14)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (26)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (34)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (37)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (39)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1677

Query Match          51.0%; Score 49; DB 10; Length 40;
Best Local Similarity 62.5%; Pred. No. 1.4;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
```

```
Oy      2 KVAKOEKKKKKTGRAK 17
        1-:|||||11-1
Db      12 KXKKKKKKKKGGRXK 27

RESULT 7
US-09-925-302-572
; Sequence 572, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 572
; LENGTH: 145
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-302-572

Query Match          51.0%; Score 49; DB 10; Length 145;
Best Local Similarity 75.0%; Pred. No. 4.6;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy      7 EKKKKKTGRAKR 18
        |||||11-11
Db      113 EKKKRRAGRSKR 124

RESULT 8
US-09-925-299-1087
; Sequence 1087, Application US/09925299
; Patent No. US2002005627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1087
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (40)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (42)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (52)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (75)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (83)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
```

```

; NAME/KEY: SITE
; LOCATION: (84)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (85)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (86)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (99)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (113)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (115)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (122)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (124)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (129)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (132)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (146)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (147)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (149)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (153)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-09-925-299-1087

Query Match          51.0%; Score 49; DB 10; Length 154;
Best Local Similarity 62.5%; Pred. No. 4.9;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy      2 KVAKEKKKKKTGRK 17
       1 ::::||||| 11:1
       14 KKKKKKKKKKGGRK 29

RESULT 9
US-10-011-582-10
; Sequence 10, Application US/10011582
; Publication No. US20030027232A1
; GENERAL INFORMATION:
; APPLICANT: DAVIS, John Beresford
; APPLICANT: GUNTHORPE, Martin
; APPLICANT: HAYES, Phillip David
; APPLICANT: KELSELL, Rosemary Elizabeth
; TITLE OF INVENTION: No. US20030027232A1e1 Compounds
; FILE REFERENCE: GP-30241
; CURRENT APPLICATION NUMBER: US/10/011,582
; PRIOR FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: UK 0109787.2
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: UK 0026951.4
; PRIOR FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
```

```

; LENGTH: 761
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
; US-10-011-582-10

Query Match          50.5%; Score 48.5; DB 9; Length 761;
Best Local Similarity 52.6%; Pred. No. 25;
Matches 10; Conservative 7; Mismatches 1; Indels 1; Gaps 1;

Qy      1 AKVAKEKKKKKTGRKR 19
       1 ::::||||| 11:1
       76 AQLAKEEQRRKK-GRKKR 93

RESULT 10
US-10-011-582-6
; Sequence 6, Application US/10011582
; Publication No. US20030027232A1
; GENERAL INFORMATION:
; APPLICANT: DAVIS, John Beresford
; APPLICANT: GUNTHORPE, Martin
; APPLICANT: HAYES, Phillip David
; APPLICANT: KELSELL, Rosemary Elizabeth
; TITLE OF INVENTION: No. US20030027232A1e1 Compounds
; FILE REFERENCE: GP-30241
; CURRENT APPLICATION NUMBER: US/10/011,582
; PRIOR FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: UK 0109787.2
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: UK 0026951.4
; PRIOR FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 790
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
; US-10-011-582-6

Query Match          50.5%; Score 48.5; DB 9; Length 790;
Best Local Similarity 52.6%; Pred. No. 26;
Matches 10; Conservative 7; Mismatches 1; Indels 1; Gaps 1;

Qy      1 AKVAKEKKKKKTGRARR 19
       1 ::::||||| 11:1
       105 AQLAKEEQRRKK-GRKKR 122

RESULT 11
US-09-864-761-39484
; Sequence 39484, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
```



```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39484
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005598.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 7.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6.7
; US-09-864-761-39484

Query Match          50.0%; Score 48; DB 10; Length 62;
Best Local Similarity 50.0%; Pred. No. 2.8;
Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY      2 KVAKOEKKKKKTGRARR 19
Db      16 KEREEREKKKKKTGKKK 33

RESULT 12
US-09-864-761-45501
; Sequence 45501, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
```

```

; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 45501
; LENGTH: 64
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008739.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.56
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.58
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.78
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.69
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.77
; US-09-864-761-45501

Query Match          50.0%; Score 48; DB 10; Length 64;
Best Local Similarity 55.6%; Pred. No. 2.9;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY      2 KVAKOEKKKKKTGRARR 19
Db      46 KEKKKEKKKKKKKKKKRK 63

RESULT 13
US-09-864-761-35659
; Sequence 35659, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
```

```

1  PRIOR APPLICATION NUMBER: GB 24263.6
2  PRIOR FILING DATE: 2000-10-04
3  PRIOR APPLICATION NUMBER: US 60/236,359
4  PRIOR FILING DATE: 2000-09-27
5  PRIOR APPLICATION NUMBER: PCT/US01/00666
6  PRIOR FILING DATE: 2001-01-30
7  PRIOR APPLICATION NUMBER: PCT/US01/00667
8  PRIOR FILING DATE: 2001-01-30
9  PRIOR APPLICATION NUMBER: PCT/US01/00664
10 PRIOR FILING DATE: 2001-01-30
11 PRIOR APPLICATION NUMBER: PCT/US01/00669
12 PRIOR FILING DATE: 2001-01-30
13 PRIOR APPLICATION NUMBER: PCT/US01/00665
14 PRIOR FILING DATE: 2001-01-30
15 PRIOR APPLICATION NUMBER: PCT/US01/00668
16 PRIOR FILING DATE: 2001-01-30
17 PRIOR APPLICATION NUMBER: PCT/US01/00663
18 PRIOR FILING DATE: 2001-01-30
19 PRIOR APPLICATION NUMBER: PCT/US01/00662
20 PRIOR FILING DATE: 2001-01-30
21 PRIOR APPLICATION NUMBER: PCT/US01/00661
22 PRIOR FILING DATE: 2001-01-30
23 PRIOR APPLICATION NUMBER: PCT/US01/00670
24 PRIOR FILING DATE: 2001-01-30
25 PRIOR APPLICATION NUMBER: US 60/234,687
26 PRIOR FILING DATE: 2000-09-21
27 PRIOR APPLICATION NUMBER: US 09/608,408
28 PRIOR FILING DATE: 2000-06-30
29 PRIOR APPLICATION NUMBER: US 09/774,203
30 PRIOR FILING DATE: 2001-01-29
31 NUMBER OF SEQ ID NOS: 49117
32 SOFTWARE: Annoncad Sequence Listing Engine vers. 1.1
33 SEQ ID NO 35659

```

```

Query Match          49.0%:  Score 47:  DB 10:  Length 80:
Best Local Similarity 50.0%:  Pred. No. 4.9:
Matches 12:  Conservative 2:  Mismatches 4:  Indels 6:  Gaps 1
OY      2  KVAKOEKKKKKTGRA-----KRR 19
      1  ::::|||||  ||
      41  KKKKKKKKKKKGGGGGGGRRRR 64
Db

RESULT 14
US-09-925-300-1667
: Sequence 1667, Application US/09925300
: Patent No. US20020151681A1
: GENERAL INFORMATION:
: APPLICANT: Craig Rosen,
: APPLICANT: Steve Ruben
: TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
: FILE REFERENCE: PA101
: CURRENT APPLICATION NUMBER: US/09/925.300
: CURRENT FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: PCT/US00/05988
: PRIOR FILING DATE: 2000-03-08
: PRIOR APPLICATION NUMBER: 60/124,270

```

```

: PRIOR FILING DATE: 1999-03-12
: NUMBER OF SEQ ID NOS: 1890
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1667
: LENGTH: 521
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-925-300-1667

```

Query Match	49.0%	Score 47	DB 10	Length 521
Best Local Similarity	56.2%	Pred. No.	28	
Matches	9	Conservative	4	Mismatches 3
				Indels 0
				Gaps 0

```
QY      3 VAKQEKKKKKTGRAKR 18
        | : : | | | | | | :
Db     154 VEEKKKKKKKKGRMKK 169
```

RESULT 15
US-09-738-626-4468

; Sequence 4468, Application US/09738626
; Publication No. US20020197605A1
; Copyright Information:

```

GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738, 626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 4468
LENGTH: 101
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-4468

```

Query Match	48.4%	Score	46.5	DB	9	Length	101
Best Local Similarity	52.4%	Pred. No.	7				
Matches	11	Conservative	5	Mismatches	4	Indels	1
						Gaps	1

```
Qy 1 AKVAKQEKKKKKTGR-AKRRA 20
    :|:| | |: | |:| |
Db 5 SKIAKNEKRKEIVARYAERRA 25
```

Search completed: February 11, 2003, 17:06:23
Job time : 28 secs


```

XX (UYLE-) RIKSUNIV LEIDEN.
PA
XX Felsma RIJ, Hiemstra PS, Nibbering PH, Pauwels EKJ.
PI Van Den Barselaar MT;
XX
XX MPI; 1999-070214/06.
DR
PT New antimicrobial peptides derived from ubiquitin - useful for
PT the prophylaxis, diagnosis and treatment of infections in humans and
PT animals
XX
XX Claim 7; Page 24; 48pp; English.
PS
XX Sequences AAW95381-389 represent antimicrobial peptide fragments derived
CC from ubiquitin. Ubiquitin or optionally modified peptide fragments
CC of ubiquitin, may be used for the treatment, diagnosis, or
CC prophylaxis of infections in humans and animals. In particular the
CC products and methods are directed against microbial infections caused by
CC pathogenic Gram-positive Staphylococcus aureus, including antibiotic
CC resistant strains, Listeria monocytogenes, and Gram-negative antibiotic
CC resistant Klebsiella pneumoniae, E. coli, enterococci, and Salmonella
CC typhimurium bacteria, micro-organisms difficult to treat such as
CC Mycobacterium avium and M. fortuitum, fungi such as Candida albicans,
CC Cryptococcus neoformans, and Aspergillus fumigatus, viruses, in
CC particular enveloped viruses, and parasites such as Trypanosoma cruzi and
CC Taxoplasma gondii.
XX
XX Sequence 20 AA:
SQ
XX
XX Query Match 100.0%; Score 96; DB 20; Length 20;
XX Best Local Similarity 100.0%; Pred. No. 9.5e-07;
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AKVAKOEKKKKTKGRAKRR 20
DB 1 AKVAKOEKKKKTKGRAKRR 20

```

```

XX New antimicrobial peptides derived from ubiquitin - useful for
PT the prophylaxis, diagnosis and treatment of infections in humans and
PT animals
XX
XX Claim 3; Page 23; 48pp; English.
PS
XX Sequences AAW95381-389 represent antimicrobial peptide fragments derived
CC from ubiquitin. Ubiquitin or optionally modified peptide fragments
CC of ubiquitin, may be used for the treatment, diagnosis, or
CC prophylaxis of infections in humans and animals. In particular the
CC products and methods are directed against microbial infections caused by
CC pathogenic Gram-positive Staphylococcus aureus, including antibiotic
CC resistant strains, Listeria monocytogenes, and Gram-negative antibiotic
CC resistant Klebsiella pneumoniae, E. coli, enterococci, and Salmonella
CC typhimurium bacteria, micro-organisms difficult to treat such as
CC Mycobacterium avium and M. fortuitum, fungi such as Candida albicans,
CC Cryptococcus neoformans, and Aspergillus fumigatus, viruses, in
CC particular enveloped viruses, and parasites such as Trypanosoma cruzi and
CC Taxoplasma gondii.
XX
XX Sequence 18 AA:
SQ
XX
XX Query Match 91.7%; Score 88; DB 20; Length 18;
XX Best Local Similarity 100.0%; Pred. No. 1.1e-05;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 KVAKOEKKKKTKGRAKR 19
DB 1 KVAKOEKKKKTKGRAKR 18

```

```

RESULT 2
AAW95385
ID AAW95385 standard; peptide; 18 AA.
XX
XX AAW95385;
AC
XX
XX
XX 17-MAR-1999 (first entry)
DT
XX
XX Antimicrobial peptide fragment from ubiquitin (residues 18-35).
DE
XX
XX ubiquitin; treatment; diagnosis; prophylaxis; infection; microbial;
XX pathogenic; Gram-positive bacteria; antimicrobial; Staphylococcus aureus;
XX Listeria monocytogenes; Gram-negative; Klebsiella pneumoniae; E. coli;
XX enterococcus; Salmonella typhimurium; Mycobacterium avium; M. fortuitum;
XX fungus; Candida albicans; Cryptococcus neoformans; Aspergillus fumigatus;
XX virus; parasite; Trypanosoma cruzi; Taxoplasma gondii.
XX
XX Synthetic.
OS
XX Mus sp.
XX
XX MO9854314-A1.
XX
XX 03-DEC-1998.
PD
XX
XX 29-MAY-1998; 98MO-NL00311.
PF
XX
XX 29-MAY-1997; 97NL-1006164.
PR
XX
XX (UYLE-) RIKSUNIV LEIDEN.
PA
XX
XX Felsma RIJ, Hiemstra PS, Nibbering PH, Pauwels EKJ.
PI Van Den Barselaar MT;
XX
XX MPI; 1999-070214/06.
DR

```

```

RESULT 3
AAW03681
ID AAW03681 standard; peptide; 41 AA.
XX
XX AAW03681;
AC
XX
XX
XX 31-DEC-1996 (first entry)
DT
XX
XX Long term potentiation inducing peptide.
DE
XX
XX Long term potentiation inducing peptide; mouse; brain; senile dementia;
XX cerebrovascular; Alzheimer's disease.
XX
XX Mus musculus.
OS
XX
XX JP08176193-A.
XX
XX 09-JUL-1996.
XX
XX 23-DEC-1994; 94JP-0336143.
XX
XX 23-DEC-1994; 94JP-0336143.
XX
XX (SOSE-) SOSEI KK.
XX
XX MPI; 1996-368226/37.
XX
XX A long-term potentiation inducing peptide - useful for the diagnosis
XX and treatment of senile dementia
XX
XX Claim 1; Page 6; 7pp; Japanese.
XX
XX This is the sequence of a long term potentiation inducing peptide
XX isolated from mice brains by centrifugation, extraction, dialysis,
XX reverse phase HPLC and gel filtration chromatography (Sephadex C-25).
XX The peptide is useful for diagnosis and treatment of senile dementia
XX e.g. cerebrovascular or Alzheimer's dementias.
XX
XX Sequence 41 AA:
SQ
XX
XX Query Match 91.7%; Score 88; DB 17; Length 41;

```

Best Local Similarity 100.0%; Pred. No. 2.3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KVAKEKKKKKTGRAKRR 19
| | | | | | | | | | | | | | | | | |
DB 18 KVAKEKKKKKTGRAKRR 35

RESULT 4

AAK48204
ID AAK48204 standard; protein: 59 AA.

AC AAK48204;

DT 12-JUL-1994 (first entry)

DE Heparin-binding protein.

KW Cell growth agent; wound; bone disease; treatment; treating; agent.

OS Rattus norvegicus.

FN JP05339287-A.

PD 21-DEC-1993.

PF 05-JUN-1992; 92JP-0145125.

PR 05-JUN-1992; 92JP-0145125.

PA (FARH) HOECHST JAPAN LTD.

DR WPI: 1994-031824/04.

PT New heparin-binding protein - used as a cell growth agent for

PT treatment of wounds and bone disease

PS Claim 1; Page 3; 4pp; Japanese.

CC The sequence is that of a heparin binding-protein which is useful as

CC a cell growth agent and in the treatment of wounds and bone disease.

SQ Sequence 59 AA;

Query Match 91.7%; Score 88; DB 15; Length 59;
Best Local Similarity 100.0%; Pred. No. 3.2e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KVAKEKKKKKTGRAKRR 19
| | | | | | | | | | | | | | | | | |
DB 18 KVAKEKKKKKTGRAKRR 35

RESULT 5

AAW95381
ID AAW95381 standard; peptide: 59 AA.

AC AAW95381;

DT 17-MAR-1999 (first entry)

DE Antimicrobial peptide fragment from ubiquicidine.

KW Ubiquicidine; treatment; diagnosis; prophylaxis; infection; microbial;

KW pathogenic; Gram-positive bacteria; antimicrobial; *Staphylococcus aureus*;

KW *Listeria monocytogenes*; Gram-negative; *Klebsiella pneumoniae*; *E. coli*;

KW *enterococcus*; *Salmonella typhimurium*; *Mycobacterium avium*; *M. fortitum*;

KW fungus; *Candida albicans*; *Cryptococcus neoformans*; *Aspergillus fumigatus*;

OS Synthetic.
OS Mus sp.
XX
XX

PN W09854314-A1.

PD 03-DEC-1998.

PF 29-MAY-1998; 98WO-NL00311.

PR 29-MAY-1997; 97NL-1006164.

PA (UYLE-) RIJKSUNIV LEIDEN.

PI Feltma RIJ, Hiemstra PS, Nibbeling PH, Pauwels EJC;

PI Van Den Barselaar MT;

DR WPI: 1999-070214/06.

PT New antimicrobial peptides derived from ubiquicidine - useful for
PT the prophylaxis, diagnosis and treatment of infections in humans and
PT animals

PS Claim 2; Page 23; 48pp; English.

CC Sequences AAW95381-389 represent antimicrobial peptide fragments derived

CC from ubiquicidine. Ubiquicidine or optionally modified peptide fragments

CC of ubiquicidine, may be used for the treatment, diagnosis, or

CC prophylaxis of infections in humans and animals. In particular the

CC products and methods are directed against microbial infections caused by

CC pathogenic Gram-positive *Staphylococcus aureus*, including antibiotic

CC resistant strains, *Listeria monocytogenes*, and Gram-negative antibiotic

CC resistant *Klebsiella pneumoniae*, *E. coli*, *enterococci*, and *Salmonella*

CC typhimurium bacteria, micro-organisms difficult to treat such as

CC *Mycobacterium avium* and *M. fortitum*, fungi such as *Candida albicans*,

CC *Cryptococcus neoformans*, and *Aspergillus fumigatus*, viruses, in

CC particular enveloped viruses, and parasites such as *Trypanosoma cruzi* and

CC *Taxoplasma gondii*.

SQ Sequence 59 AA;

Query Match 91.7%; Score 88; DB 20; Length 59;
Best Local Similarity 100.0%; Pred. No. 3.2e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KVAKEKKKKKTGRAKRR 19
| | | | | | | | | | | | | | | | | |
DB 18 KVAKEKKKKKTGRAKRR 35

RESULT 6
ABP42408
ID ABP42408 standard; Protein: 66 AA.

AC ABP42408;

DT 22-AUG-2002 (first entry)

DE Human ovarian antigen HOCW79, SEQ ID NO:3540.

KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;

KW ovarian cancer; breast cancer; tumour; reproductive system disorder;

KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;

KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;

KW inflammatory condition; immune disorder; blood disorder;

KW cardiovascular disorder; respiratory disorder; neurological disorder;

KW gastrointestinal disorder; urinary system disorder; drug screening;

KW gene therapy; chromosome mapping; forensic analysis;

KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;

KW antiinflammatory; gynaecological; reproductive.

OS Homo sapiens.
OS W0200200677-A1.
OS 03-JAN-2002.
XX
XX
XX
XX

PF 07-JUN-2001; 2001WO-US18569.
XX
XX 07-JUN-2000; 2000US-209467P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
DR WPI: 2002-147878/19.
XX
DR N-PSDB; AB055485.
XX
PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g.
PT ovarian cancer), immune disorders, cardiovascular disorders and
PT neurological diseases -
XX
PS Claim 11: SEQ ID NO 3540; 2922pp; English.
XX
XX The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (AB054131-AB056305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents a human ovarian antigen of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 66 AA:
Query Match 91.7%; Score 88; DB 23; Length 66;
Best Local Similarity 100.0%; Pred. No. 3.6e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KVAKOEKKKKKTGRARR 19
DB 25 KVAKOEKKKKKTGRARR 42

RESULT 7
AAG75965
ID AAG75965 standard; Protein: 157 AA.
XX
XX AAG75965;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:6729.
XX
XX Human: colon cancer; colon cancer antigen; diagnosis; detection;
KM colorectal carcinoma; chromosome 11.
XX
OS Homo sapiens.

XX
XX WO200122920-A2.
PN
XX
XX 05-APR-2001.
PD
XX
XX 28-SEP-2000; 2000WO-US26524.
PF
XX
XX 29-SEP-1999; 99US-0157137.
PR
XX 03-NOV-1999; 99US-0163280.
XX
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
PI WPI: 2001-235357/24.
XX
DR N-PSDB; AAH35370.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 11: Page 8192-8193; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77768 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB7789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 157 AA:
Query Match 91.7%; Score 88; DB 22; Length 157;
Best Local Similarity 100.0%; Pred. No. 7.9e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KVAKOEKKKKKTGRARR 19
DB 123 KVAKOEKKKKKTGRARR 140

RESULT 8
AAB43549
ID AAB43549 standard; Protein: 170 AA.
XX
XX AAB43549;
XX
DT 08-FEB-2001 (first entry)
XX
XX Human cancer associated protein sequence SEQ ID NO:994.
XX
DE Human cancer associated protein sequence SEQ ID NO:994.
XX
XX Human: cancer associated gene; cancer antigen; detection; cancer;
KM diagnosis; cytostatic; proliferative; vulnerrary; immunomodulator;
XX antidiabetic; antiasclmatic; antitumouric; antiarthritic; antiviral;
XX antiinflammatory; antilymphoid; antiallergic; antibacterial; cardiac;
KM dermatological; neuroprotective; thrombolytic; coagulant; neotropic;
XX vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KM immune disorder; haematopoietic cell disorder; autoimmune disorder;
XX allergic reaction; graft versus host disease; organ rejection;
KM haemostatic; thrombolytic; cardiovascular disorder; infection;
KM neurological disease; drug screening.

XX OS Homo sapiens.
XX PN WO200055350-A1.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000; 2000WO-US05882.
XX PR 12-MAR-1999; 99US-0124270.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX DR WPI: 2000-587533/55.
XX DR N-PSDB: AAC77758.
XX PT Novel isolated nucleic acids comprising sequences encoding peptides
XX PT useful for treating or diagnosing e.g. cancer -
XX PS Claim 11; Page 1568-1569; 2352pp; English.
XX CC AAC77607 to AAC78448 encode the human cancer associated proteins given
XX CC in AAB43398 to AAB44239. The proteins can have activities based on the
XX CC tissues and cells the genes are expressed in. Example of activities
XX CC include: cytostatic; proliferative; vulnerary; immunomodulator;
XX CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
XX CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
XX CC neurotropic; vasotropic; antipsoriatic and antiangiogenic. The
XX CC polynucleotides and polypeptides can be used for preventing, treating or
XX CC ameliorating medical conditions and diagnosing pathological conditions.
XX CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
XX CC the present invention may be used to treat immune disorders by activating
XX CC or inhibiting the proliferation, differentiation or mobilisation of
XX CC immune cells, to treat disorders of haematopoietic cells, autoimmune
XX CC disorders, allergic reactions, graft versus host disease and organ
XX CC rejection, modulate haemostatic or thrombolytic activity, modulate
XX CC inflammation, cancers, cardiovascular disorders, neurological disease and
XX CC bacterial or viral infections. The peptides, nucleotides, antibodies, and
XX CC agonists and antagonists may be also be used in drug screens. AAC78449 to
XX CC AAC78457 and AAB44240 represent sequences used in the exemplification of
XX CC the present invention.
XX SQ Sequence 170 AA:
Query Match 91.7%; Score 88; DB 21; Length 170;
Best Local Similarity 100.0%; Pred. No. 8.5e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 KYAKOEKKKKTKGRAKRR 19
DB 129 KYAKOEKKKKTKGRAKRR 146
RESULT 9
ABB70963
ID ABB70963 standard; Protein; 132 AA.
XX ABB70963;
XX AC
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 39681.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX XX

PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI: 2001-656860/75.
XX DR N-PSDB: ABL15066.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Disclosure; SEQ ID NO 39681; 21pp + sequence listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABBS7737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 132 AA:
Query Match 86.5%; Score 83; DB 22; Length 132;
Best Local Similarity 94.4%; Pred. No. 0.00033;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 KYAKOEKKKKTKGRAKRR 19
DB 90 KYAKOEKKKKTKGRAKRR 107
RESULT 10
AAG01569
ID AAG01569 standard; Protein; 130 AA.
XX AAG01569;
XX AC
XX DT 06-OCT-2000 (first entry)
XX DE Human secreted protein, SEQ ID NO: 5650.
XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping.
XX OS Homo sapiens.
XX PN EP1033401-A2.
XX PD 06-SEP-2000.
XX PF 21-FEB-2000; 2000EP-0200610.
XX PR 26-FEB-1999; 99US-0122487.
XX PA (GENSET) GENSET.
XX PI Dumas Mline Edwards J, Duclert A, Giordano J;
XX DR WPI: 2000-500381/45.
XX DR N-PSDB: AAC01575.
XX XX

CC and gene mapping, and in recombinant production of (II). The

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG3037 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 108 AA:

Query Match 82.3%; Score 79; DB 22; Length 108;
 Best Local Similarity 100.0%; Pred. No. 0.00098;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 AKQEKKKKTGRAKRR 19
 |||||
 DB 69 AKQEKKKKTGRAKRR 84

RESULT 13

AAU30872
 ID AAU30872 standard; Protein: 312 AA.

XX AAU30872;

DT 18-DEC-2001 (first entry)

XX Novel human secreted protein #1363.

XX Human; vaccination: gene therapy; nutritional supplement;

KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;

KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX Homo sapiens.

XX WO200179449-A2.

XX 25-OCT-2001.

XX 16-APR-2001; 2001WO-US08656.

XX 18-APR-2000; 2000US-0552929.

XX 26-JAN-2001; 2001US-0770160.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI: 2001-611725/70.

XX Nucleic acids encoding a range of human polypeptides, useful in genetic

XX vaccination, testing and therapy -

XX Claim 20; Page 364; 765pp; English.

CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising
 CC the nucleic acids encoding the polypeptides and cells genetically
 CC engineered to express them are also useful for producing the proteins.

CC The proteins are useful in genetic vaccination, testing and
 CC therapy, and can be used as nutritional supplements. They may be used to
 CC increase stem cell proliferation; to regulate haematopoiesis; and in
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and
 CC in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid
 CC sequences of novel human secreted proteins of the invention.

XX Sequence 312 AA:

Query Match 75.5%; Score 72.5; DB 22; Length 312;
 Best Local Similarity 89.5%; Pred. No. 0.02;
 Matches 17; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 2 KVAKQ-EKKKKKTGRAKRR 19
 |||||
 DB 66 KVAKQEKKKKTGRAKRR 84

RESULT 14

AAO08731
 ID AAO08731 standard; Protein: 40 AA.

XX AAO08731;

DT 06-NOV-2001 (first entry)

XX Human polypeptide SEQ ID NO 22623.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;

XX nervous system disorders; arthritis; inflammation.

XX Homo sapiens.

XX WO200164835-A2.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

XX 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI: 2001-514838/56.

XX N-PSDB; AA188662.

XX Isolated nucleic acids and polypeptides, useful for preventing

XX diagnosing and treating e.g. leukaemia, inflammation and immune

XX disorders -

XX Claim 20; SEQ ID NO 22623; 1399pp + Sequence Listing; English.

CC The invention relates to human polynucleotides (AA179941-AA193841) and
 CC the encoded proteins (AAO0010-AAO13910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

